SIRE OF FETUS EFFECTS AND ENVIRONMENTAL CORRELATIONS FOR DAIRY CATTLE PRODUCTION AND REPRODUCTION

BY

JOAQUIM MOYA

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1985

ACKNOWLEDGMENTS

I want to thank most sincerely Dr. Charles J. Wilcox, chairman of my committee and major advisor, for his invaluable guidance during the development of this research and throughout my Ph.D. program. Dr. Wilcox has the rare ability to summarize in a simple manner concepts that sometimes tend to be obscurely complicated by other people. I highly appreciate Dr. Wilcox's guidance, patience, and the friendship that was available at all times.

I also want to thank the members of my committee, Dr. R. J. Collier, Dr. F. G. Martin, and Dr. J. R. Simpson, for reviewing this manuscript and giving me valuable suggestions. Dr. Martin's meticulous work reviewing the mathematical proofs of Chapter VI and his good ideas and support are especially appreciated. Dr. Simpson's input during not only the research but also throughout my program was very valuable. Dr. Collier's guidance and friendship contributed to successful completion of my program.

During the last year of my Ph.D. program, I had a very rewarding relationship with Dr. M. D. DeLorenzo. As a result of long conversations with him, I was able to

extract valuable ideas. His work reviewing this manuscript is highly appreciated also.

The help of Mrs. Marilyn K. Tucker, scientific programmer in IFAS-statistics, in finding solutions to computing aspects of data management was very valuable, and I am indebted to her for this help.

My relationship with Dr. Thatcher, pioneer researcher of physiological aspects of size of fetus effects in dairy cattle, has been consistently rewarding. I highly appreciate the human touch that I always found in him.

Finally, special recognition is given to Sharon, the English teacher always fighting Spanish windmills trying to improve my English, to Sharon, the lovely wife, and to Nehmi and Nuria, my daughters, for all their moral support.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS.....ii

LIST OF TABLES.....vi

LIST OF FIGURES.....viii

ABBREVIATIONS.....ix

ABSTRACT.....x

CHAPTERS

I INTRODUCTION.....1

II LITERATURE REVIEW.....4

Sire of Fetus Effects.....4
Dairy Breed Sires.....4
Beef Breed Sires.....20
Summary.....21
Environmental Correlations.....22
Statistical Aspects of Variance Component Estimation.....34
ANOVA Approach.....35
MINQUE and MIVQUE Approach.....38

III SIRE OF FETUS EFFECTS.....41

Introduction.....41
Material and Methods.....44
Data.....44
Analyses.....48
Model.....49
Henderson Method 1.....51
Henderson Method 3.....51
MIVQUEO.....53
Results and Discussion.....54
Summary and Conclusions.....62

IV ENVIRONMENTAL CORRELATIONS.....65

Introduction.....65
Material and Methods.....70
Data.....70
Analyses.....70
Henderson Method 3.....71
MIVQUEO.....71
Results and Discussion.....72
Summary and Conclusions.....81

V COMPUTING EFFICIENCY OF HENDERSON METHOD 3 AND MIVQUEO.....83

Introduction.....83
Material and Methods.....84
Results and Discussion.....84
Summary and Conclusions.....87

VI EFFECTS OF MISIDENTIFICATION OF SIRE OF FETUS ON ESTIMATES OF SIRE OF FETUS VARIANCE.....88

Introduction.....88
Material and Methods.....89
Results and Discussion.....90
Summary and Conclusions.....95

VII SUMMARY AND CONCLUSIONS.....96

APPENDIX PROOFS FOR FORMULAS IN TABLES 6.2 AND 6.3....100

BIBLIOGRAPHY.....104

BIOGRAPHICAL SKETCH.....113

LIST OF TABLES

Table		Page
2.1	Components of Variance for Paternal Half-Sibs	28
2.2	Environmental Correlations (c^2) Between Paternal Half-Sisters	31
3.1	Existing Research Concerning SOF Effects on Milk Production	43
3.2	Data Sets, Distributed by State and Breed	46
3.3	Jersey Data Set, Means and Standard Deviations	48
3.4	Jersey Data Set by State	49
3.5	Type I SS and RSS for Main Effects in a Random Model	52
3.6	Henderson Method 1 Variance Component Estimates for Complete Model, Allowing Negative Estimates	54
3.7	Henderson Method 1 Variance Component Estimates for Complete Model, Setting Negative Estimates Equal to Zero	55
3.8	MIVQUEO for SOF, Random Complete Model	57
3.9	Henderson Method 3 Variance Component Estimates for SOF, Model with Only Main Effects	58
3.10	MIVQUEO for SOF, Mixed Model with Only Main Effects	58
3.11	Henderson Method 3 Variance Component Estimates for SOF After Fitting HYS, SOC and HYS x SOC Interaction	60
3.12	Henderson Method 3 Variance Component Estimates for SOF Confounded with HYS x SOF, SOC x SOF and HYS x SOC x SOF Interactions	62

4.1 Existing Research Concerning Environmental Correlation.....68

4.2 Type I SS and RSS for a Random Model.....72

4.3 Henderson Method 3 Variance Components Estimates for First Order Interactions.....74

4.4 MIVQUEO for First Order Interactions.....75

4.5 Henderson Method 3 Variance Component Estimates for HYS x SOC.....76

4.6 Henderson Method 3 Variance Component Estimates for HYS x SOF.....76

4.7 Henderson Method 3 Variance Component Estimates for SOC x SOF.....77

4.8 MIVQUEO for HYS x SOC.....77

4.9 MIVQUEO for HYS x SOF.....78

4.10 MIVQUEO for SOC x SOF.....78

5.1 Computing Time for Method 3 and MIVQUEO Estimating Variance Components of Main Effects....85

5.2 Computing Time for Method 3 and MIVQUEO Estimating Variance Components of First Order Interactions.....86

6.1 ANOVA Table to Study Effects of SOF Misidentification on Estimates of Variance Components.....91

6.2 Sums of Squares for Models with SOF Misidentification.....92

6.3 Expectations of SOF and Residual Components of Variance for Models with SOF Misidentification....94

LIST OF FIGURES

Figure		Page
2.1	Fetal Sire Effects on Dam Performance, Model by Skjervold and Fimland	5
2.2	Fetal Sire Effects on Dam Performance, Model by Van Vleck	15

ABBREVIATIONS

SOF:	Sire of fetus
SOC:	Sire of cow
HYS:	Herd-Year-Season
MILKYD:	Milk yield
FATYD:	Fat yield
FAT%:	Fat %
DAYSOP:	Days open
GESTL:	Gestation length
ME:	Mature equivalent
AI:	Artificial insemination
DHI:	Dairy Herd Improvement
MIVQUEO:	Minimum Variance Quadratic Unbiased Estimator with priors zero
MINQUE:	Minimum Norm Quadratic Estimation
REML:	Restricted Maximum Likelihood
c^2:	Environmental correlations

Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

SIRE OF FETUS EFFECTS AND ENVIRONMENTAL CORRELATIONS FOR DAIRY CATTLE PRODUCTION AND REPRODUCTION

BY

JOAQUIM MOYA

May 1985

Chairman: Charles J. Wilcox
Major Department: Animal Science

After screening a data set consisting of 2.2 million individual lactation records from cows in NC, SC, GA and FL, a subset of 14,147 records from Jersey cows was obtained. Henderson Methods 1 and 3, and MIVQUEO techniques, were used to obtain estimates of HYS (herd-year-season), SOC (sire of cow) and SOF (sire of fetus) variance components. With Method 1 SOF effects estimates were moderately large (9 to 15%) for MILKYD (milk yield), FATYD (fat yield), FAT% (fat %) and GESTL (gestation length) and very small for DAYSOP (days open). With Method 3 SOF effects were considerably smaller (<1 to 3%) for production and reproduction traits, except for GESTL which was moderate (10%). MIVQUEO estimates from mixed models were practically identical to those obtained with Method 3.

Henderson Method 3 and MIVQUEO with mixed models techniques were used to estimate variance composition of the first order interactions HYS x SOC, HYS x SOF and SOC SOF for MILKYD, FATYD, FAT%, DAYSOP and GESTL. Arithmetic and weighted means of the Henderson Method 3 estimates by state, expressed as percentages of the total variance due to interaction, were small and positive (0.4 to 6.5%). Means of MIVQUEO were similar, but they tended to be slightly smaller and some were negative (-0.8 to 3.7%). These interactions suggest existence of covariances betwee HYS, SOC and SOF, therefore, their summation is a good description of the common environmental correlation c^2. Henderson Method 3 estimates for c^2 following this approac were 10.1, 9.4, 13.2, 7.3 and 10.2% for MILKYD, FATYD, FAT%, DAYSOP and GESTL. MIVQUEO were 2.6, 2.5, 4.0, 2.1 and 4.1% for the same traits. Results obtained in this research using Henderson Method 3 and MIVQUEO, in agreement with results obtained by other authors using Method 1, indicated that the value of 0.14 for c^2 presently used by USDA to evaluate sires is biased slightly upwards. Results also indicated that MIVQUEO is more efficient than Method 3 in use of computer time and its efficiency increases with the level of total time used.

Effects of SOF misidentification on estimates of SOF variance components were studied using Henderson Method 1. It was found that expected values for error term

CHAPTER I
INTRODUCTION

In the mid-seventies, Skjervold and Fimland (76) in Norway and Adkinson et al. (2) in Florida investigated the amount of total variance for milk production that can be attributed to SOF effects. With different approaches, but both using Henderson Method 1 (30), they had different results. The Norwegian researchers found that SOF was responsible for approximately 1% of the total variation, and the Florida researchers found that SOF accounted for approximately 10% of the total variation for milk production.

In 1979, Johnson and Van Vleck (36, 37) from Cornell used Henderson Method 1 and Method 3 (30) to estimate the variance component of SOF effects. They found that when Method 3 was used, the estimates were smaller than those obtained with Method 1. Attempts to use MINQUE techniques to estimate variance components for SOF were made by Wickham (92) at Cornell and Sharma et al. (74) in Florida.

The advantages of Method 3 over Method 1 in estimation of variance components for unbalanced data sets with mixed models have been well known since 1953 when Henderson introduced his variance component estimation procedures

(30). Method 3 estimates are unbiased by the existence of covariances between main effects. Method 1 estimates are biased if covariances exist. Unfortunately, Method 3 is very demanding in computer time and until recently, when larger and faster computers have been available, Method 3 has not been widely used in estimation of variance components with large data sets.

A new approach to estimate variance components, called MIVQUEO, was proposed by Hartley et al. (26) in 1978. This technique is not based on traditional analysis of variance procedures and is supposed to be more efficient in the use of computer time.

The main goal of our research was twofold. In a first step, results were compared using Henderson Method 1, Method 3, and MIVQUEO, with random and mixed models to estimate variance components for the main effects HYS, SOC and SOF, emphasizing the role of SOF effect on the total variance.

In a second step, variance components of the interactions among main effects were estimated using the same techniques listed above. The role of those interactions in the constitution of the environmental correlation c^2 was explored. Several researchers (3, 11, 32, 33, 39, 40, 47) have identified c^2 as the interaction HYS x SOC, but the role of the interactions of SOF with the other main effects on c^2 has not been studied.

Secondary aspects of research included the use of computer time during the estimation of variance components with Henderson Method 3 and MIVQUEO, and a comparison of the efficiency of each of these methods in the use of computing resources. Additionally, the effect of SOF misidentification on estimates of variances also was studied.

Most analyses in this research were performed using a subset of Jersey records. The main reason for this was that the number of observations in this data set was adequate for the type of analyses carried out. The use of the considerably larger subset of Holstein records would have made the cost of the analyses unaffordable.

CHAPTER II
LITERATURE REVIEW

Sire of Fetus Effects

Dairy Breed Sires

As early as 1915 Gaines (19) hypothesized that SOF, through the fetus, could affect milk yield of the dam. Nevertheless, Skjervold and Fimland (76) in 1975 published the first paper supported by statistical analysis on a large data set from Norway. Their results were confirmed by Skjervold (75) in 1979. The genetic model proposed by Skjervold and Fimland (76) attempting to explain fetal sire effects on dam performance (i.e., milk production) is shown in Figure 2.1.

Neither the phenotypic correlation between fetus and dam nor heritability of fetal traits in question can be obtained directly. However, the correlation between phenotype and SOF genotype for milk production can be estimated from

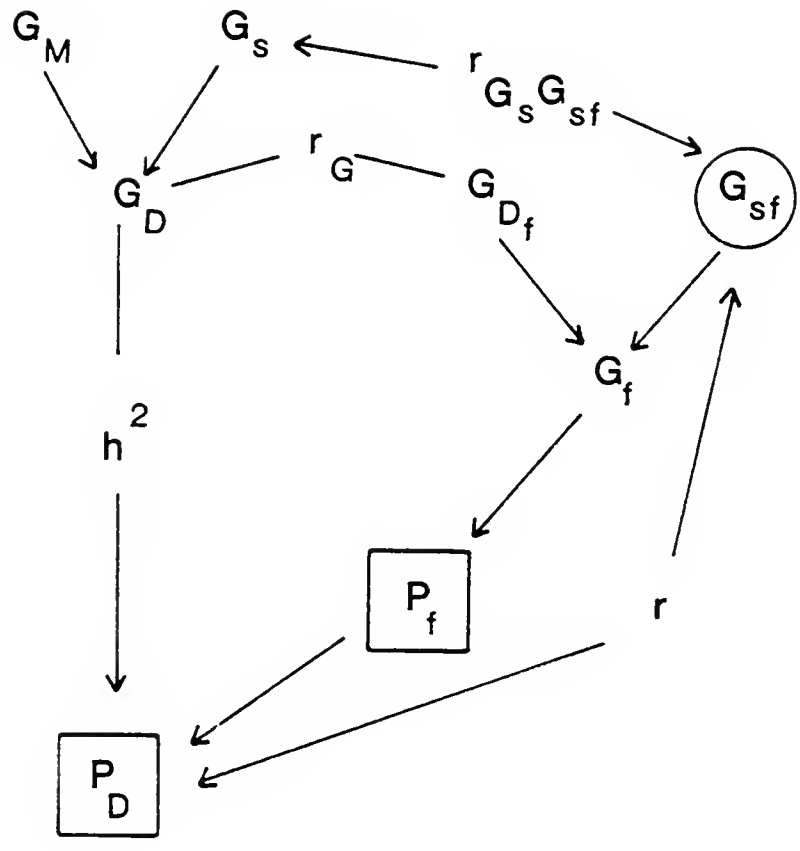
$$Pd = \mu + 1/2\ Gsf + 1/2\ Gs + R, \qquad [2.1]$$

where

Pd = phenotype for milk production of the dam,

Gsf = genotype for milk production of the SOF,

Gs = genotype for milk production of the SCC,



P_D : Phenotype of milk production

P_f : Phenotype of hormone activity of fetus

G_{sf} : Genotype of the sire of the fetus

Figure 2.1: Fetal Sire Effects on Dam Performance, Model by Skjervold and Fimland (76)

R = residual, and

$Cov(Gsf,Pd) = 1/2\ \sigma^2 Gsf$,

$V(Gsf) = \sigma^2 Gsf$, and

$V(Pd) = \sigma^2_{Pd} = 1/4\ \sigma^2 Gsf + 1/4\ \sigma^2 Gs + \sigma^2 R$.

The phenotypic correlation coefficient is estimated by

$$r_{Pd \cdot Gsf} = \frac{Cov(Gsf, Pd)}{(V(Gsf)V(Pd))^{1/2}} = \frac{1/2\ \sigma^2_{Gsf}}{\sigma_{Gsf}\sigma_{Pd}} = \frac{1/2\ \sigma_{Gsf}}{\sigma_{Pd}}$$

Using this genetic model Skjervold and Fimland (76) analyzed data from first lactation cows calving during fall to avoid effects of previous lactation and season. They were interested in lactation yields, measured as deviations from herd average, and maximum daily yields, corrected for age at calving and herd average. They analyzed 69,501 first lactation records. Mathematical models were assumed to be completely random.

Variance components for SOF effects were obtained by nesting SOC within SOF groups. The data then also were analyzed by nesting SOF within SOC groups. The combination of nested analyses gave the cross-classified component of variance for SOF and SOC. They did not find significant differences between the cross-classified components and estimates obtained in the first analysis. This finding suggested that the interaction SOC x SOF did not exist.

Skjervold and Fimland (76) presented results of these analyses by years and by nonprogeny tested sires as well as all sires. Means of correlations were 0.096 for milk yield during first lactation and 0.076 for maximum daily yield. These investigations did not detect any significant correlation between breeding value of SOC and effects of SOF.

Adkinson (1) and Adkinson et al. (2) published the first estimates of SOF effects from a large data set from the U.S.A. The initial data set for this study consisted of 945,649 lactation records from the lifetime production file for cows on Dairy Herd Improvement (DHI) programs in Southeastern U.S. After extensive screening, the data set was reduce to 3,731 Jersey and 27,200 Holstein records.

Adkinson et al. (2) used random and mixed models, including nested and cross-classified effects in various combinations to estimate SOF effects, genetic parameters and relationships of milk yield, fat yield and days open. Variance and covariance components for cross-classified effects were estimated by setting mean squares of nested effects equal to their expectations in terms of cross-classified components, and solving the resulting equations. Estimates corresponded to those resulting from Henderson Method 1 (30).

In their first efforts to determine if SOF influenced milk or fat yields of the cow following birth of the fetus,

Adkinson et al. (2) used the model

$$Y_{ijk\ell} = \mu + a_i + b_{ij} + F_k + e_{ijk\ell}, \qquad [2.2]$$

where

$Y_{ijk\ell}$ = observed value of dependent variables,

μ = fixed effect common to all records,

a_i = random effect due to i th HYS,

b_{ij} = random effect of j th SOF in the i th HYS,

F_k = fixed effects due to age, length of record, days open during the record and days dry prior to the record, all fitted to the cubic degree polynomial,

$e_{ijk\ell}$ = error, random and normally distributed.

Estimates of SOF components represented 3.01 and 0.22% of the total variance associated with random components in the model.

Since sires of cows were ignored, assortative mating could be a possible source of bias with this model. If better bulls were bred to better cows, this might result in differences between SOF groups being attributed to SOC. The following model was proposed by Adkinson et al. (2) to reduce bias from this source

$$Y_{ijk\ell} = \mu + a_i + b_{ij} + c_{ijk} + e_{ijk\ell}, \qquad [2.3]$$

where

$Y_{ijk\ell}$, μ, a_i and $e_{ijk\ell}$ are as [2.2] and

b_{ij} = random effect due to j th SOC in the i th HYS,

c_{ijk} = random effect due to k th SOF in j th SCC in i th HYS.

Dependent variables were ME milk and fat yields and days open. These analyses resulted in unreasonably large estimates for SOF components (6.00 to 13.87%). SOF components for days open were small to moderate in most of the cases (0 to 3.16%).

A within cow analysis was attempted by Adkinson et al. (2) using deviation of ME yields from adjusted herdmate averages as response variables with the following model:

$$Y_{ijk} = \mu + a_i + b_{ij} + e_{ijk} , \qquad [2.4]$$

where

Y_{ijk} = observed value of dependent variables,

μ = fixed effect common to all records,

a_i = random effect of i th cow,

b_{ij} = random effect of j th SOF in i th cow.

Results did not agree closely with previous analyses. SOF effects estimates for Jerseys were even higher (21.13 and 12.37%) for herdmate deviated milk and fat yields but were zero for Holsteins. The model described in [2.3] for analyses of ME records then was applied to herdmade deviated records. Estimates of SOF effects were not nearly as extreme in this analysis and agreed with those for ME records.

Adkinson et al. (2) concluded that inclusion of SOF in current programs of sire proving was indicated, when computationally feasible, if SOF effects were real and measurable. Due to the apparent absence of interactions of SOC and SOF, they suggested that present sire proofs were unbiased by SOF as long as mating of sires to daughters of other sires was random. However, future considerations of SOF, perhaps by adjustment of milk records for SOF, could result in worthwhile reductions in error variances and increased repeatability of proofs and heritabilities.

Auran et al. (5) in 1977 investigated effects of concurrent sires, i.e., sire of the fetus that the cow carried during lactation, on milk production initiated by previous calving. Using monthly and cumulative first lactation Norwegian records, the correlation between the phenotype of milk production of the dam and the genotype of the sire of the calf was estimated following Skjervold and Fimland's (76) methods. They used a total of 69,000 records of cows mated with 519 sires. The concurrent sire of fetus effect on milk production increased from 0.02 130 days before calving up to 0.10 60 days before calving.

Taylor et al. (78, 79) reported an experiment designed to estimate differences between high contemporary comparison (CC) sires and low CC sires on their mates milk yield. The same design was equally efficient for estimating sire differences. This experiment comprised 175

experimental lactations from mates of 16 proven bulls. Design of this experiment was a cross-over in which cows were allocated at random either to high CC bulls for the first pregnancy and low CC bulls for their second pregnancy or vice versa.

Total lactation milk was the dependent variable in least squares analysis. Estimates of mate effect on improved contemporary comparison (ICC) were obtained by regression. The regression constants (intercepts) were variable and none differed significantly from zero. The overall unweighted value was 8 ± 50 and the weighted value, using the reciprocals of the error variances as weighting factors, was -3 ± 52. The coefficients were reestimated to force the regressions through the origin. For first lactation only, the weighted regression was -0.47 ± 0.25 kg. of milk per kg. ICC, whereas for second lactations it was -0.32 ± 0.23. Breeds did not differ significantly in their regressions. Overall correlations between ICC and mate effects were -0.52 (unweighted) and -0.46 (weighted). Taylor et al. (79) concluded that, if these results were confirmed, the overall regression suggested that a bull with an ICC of +100 kg. would on average reduce each mate's milk yield by 38 ± 16 kg. (unweighted) or 33 ± 16 kg. (weighted) or by about one-third of his estimated breeding value. These results implied that as selection for higher CC bulls proceeds, potentially high-yielding

daughters produced may tend to have their yields lowered by having high CC bulls as mates.

Production and reproduction records made over 23 years at the Dairy Research Unit of the University of Florida were analyzed by Quesnel (57) and Quesnel et al. (58). The data set for this research consisted of 3,776 lactation records. Several mathematical models were used in this research in a series of least squares analyses to estimate SOF effects, parturition sire effects, two previous sire effects, successful sire effects, genetic parameters and relationships between the responses measured. As dependent variables the classical production responses of yields and percentages were considered as well as reproduction responses.

Analyses were performed using Henderson Method 3 (30). The mathematical model used in the analysis of the pooled data (all breeds) was

$$Y_{ijk\ell} = \mu + a_i + b_{ij} + F_k + e_{ijk\ell} , \qquad [2.5]$$

where

$Y_{ijk\ell}$ = Response for each dependent variable,

μ = Overall mean,

a_i = SOC-breed-year-season effects,

b_{ij} = SOF effect nested in former effects,

F_k = a set of fixed effects, including

a) Continuous: age of cow, length of record, days pregnant, birth weight of

calf and gestation length (all linear, quadratic and cubic)

b) Discrete: sex of calf, horn of pregnancy, metritis, dystocia and retained placenta,

$e_{ijk\ell}$ = random and normally distributed error.

All results were expressed as a ratio of the SOF variance to the sum of the error and SOF variance.

SOF effects on the concurrent lactation were found to be nil. Quesnel (57) and Quesnel et al. (58) were aware that assortative mating could be a source of bias in estimating SOF effects on subsequent lactation. To account for this possibility, they nested SOF in SOC-cow-breed-year-season categories. Using this model, SOF effects for milk yield and fat yield were responsible for 16.2% and 12.4% of the variance. SOF effects were higher for component yields than for component percentages and smaller for reproduction traits. The authors (57, 58) concluded that there was ample evidence that SOF affected subsequent production and reproduction of the dairy cow.

SOF effects on second and third subsequent lactations also were detected by Quesnel (57) investigating carryover effects of SOF on additional lactations. Nevertheless, he was careful with the interpretation of these results given the relatively small data set used and the very large number of sire x cow cells, which left few degrees of freedom for error.

Van Vleck (87, 88) proposed a genetic model to explain fetal effects on traits of the dam somewhat differently from that postulated by Skjervold and Fimland (76). This model is in Figure 2.2.

Any trait influenced by the mate of the female (SOF) can be described by such a model. The model can be written as

$$Px = Gx + Fw + Ex + Ew$$

$$Py = Gy + Fz + Ey + Ez$$

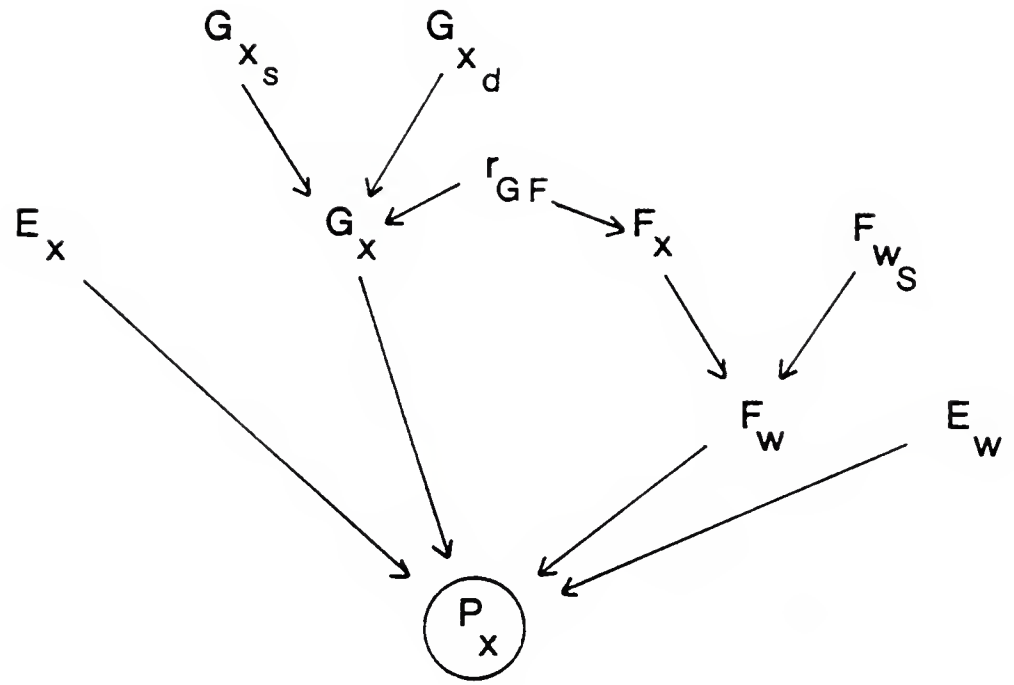
where the G's are genetic values for the direct effects on P, the F's are the genetic effects of the fetus on P, and E's are the corresponding environmental effects. In the classical model P = G + E, all effects except G would be included in E. The F effects are environmental to the animal making the record but are genetically determined. If only additive genetic effects are considered important, the convariance between Px and Py can be written as

$$Cov(Px,Py) = a_{xy}\sigma^2_G + a_{wz}\sigma^2_F + (a_{xz} + a_{yw})\ \sigma^2_{GF}$$

where σ^2_G is the variance of direct additive genetic effects, σ^2_F is the variance of additive fetal genetic effects and σ^2_{GF} is the covariance between them.

The typical random model for estimating direct and fetal sire effects and for predicting sire values is

$$P_{ijk} = s_i + t_j + (st)_{ij} + r_{ijk}$$



G_x : Direct genetic effects

E_x : Direct environmental effects

F_w : Fetal genetic effects

E_w : Fetal environmental effects

Figure 2.2: Fetal Sire Effects on Dam Performance, Model by Van Vleck (87)

where P_{ijk} is the record of the female, s_i is the effect of SOC which often is considered to be one-half the direct additive genetic value of sire i, and t_j is the effect of the SOF which is one-half the fetal additive genetic value of the SOF j. Nevertheless, Van Vleck (87) showed that

$$\sigma_s^2 = 1/4\ \sigma_G^2 + 1/16\ \sigma_F^2 + 1/4\ \sigma_{GF}^2$$

which amplified the classic concept of s being one-half the direct additive genetic value of sire i, and $\sigma_t^2 = 1/4\ \sigma_F^2$ and, finally, $\sigma_{st}^2 = 0$. This meant, as stated by Van Vleck (87), that the interaction component SOC x SOF contained no additive genetic terms, and thus would not be expected to be large. This agreed with results obtained by Skjervold and Fimland (76) and Adkinson et al. (2).

Van Vleck (87) examined the relationships s =1/2 G + 1/4 F and t = 1/2 F, and concluded that selection as now performed for s was actually selection for one-fourth of the fetal genetic value of the sire as well as the expected one-half of the direct genetic value of the sire since the sire of the cow is also the maternal grandsire of the fetus. Finally, Van Vleck (87) elaborated on the possibility of bias in evaluating sires if fetal sires were not used randomly on daughters of all sires.

Components of variance associated with SOF for milk yield and reproductive traits were studied by Johnson and Van Vleck (37). Two reproductive intervals and three

measures of milk yield were taken into consideration. The data consisted of 1.9 million second lactation records of Holstein cows enrolled in the DHI program; after screening the data set was reduced to 47,737 records.

The mathematical model applied was

$$Y_{ijk\ell} = \mu + h_i + s_j + (hs)_{ij} + ss_k + e_{ijk\ell}, \quad [2.6]$$

where

$Y_{ijk\ell}$ = dependent variables (milk yield and reproductive traits),

μ = constant,

h_i = random i th HYS effect,

s_j = random j th SOC effect,

$(hs)_{ij}$ = random HYS x SOC interaction effect,

ss_k = random k th service sire effect,

$e_{ijk\ell}$ = residual error effect.

In different analyses the service sire was either the SOF (SS1), sire of calf whose birth initiated the cows second lactation, or the sire that the cow was bred to during the second lactation (SS2), i.e., concurrent sire.

For each of the two types of service sire, the variance components were estimated using Henderson Method 1 and Method 3 (30). Using Method 1 the SS1 effect counted for 4.8 and 2.7 of the variance for 305 ME in (36) and (37) respectively. The variance component estimated for SS2 effect was larger than for SS1 effect. This surprising result was explained by the author as an associative source

of variance that appeared in the SS2 component as a result of a selectively chosen service sire after her first and early second lactation milk yields.

When Method 3 was applied to the same data (37, 38), the amount of variance accounted for by SOF was on the order of 1% for the three milk yield traits. The reduction was explained by Johnson and Van Vleck (37) as the result of the ability of Henderson Method 3 (30) to eliminate effects of nonrandom use of service sire across herds and sires of cows. The authors (37) suggested the need for investigation of non-random use of service sires as it affects current methods of dairy sire evaluation.

Hayes et al. (29), working with data from 8,817 Holstein cows enrolled in the Quebec Dairy Herd Analysis Service, investigated associations between SOF and production traits. Using mixed models and Henderson Method 3 (30), results similar to those reported by Johnson and Van Vleck (37) were found. Percentages of the variance accounted for by SOF were 1.0% for milk yield, 0.25% for fat yield, 0.5% for protein yield, 0.2% for fat percent and 0.25% for protein percent. Results were similar regardless of whether calving difficulty was included in the model.

In separate studies Wickham (92), as cited by Johnson and Van Vleck (37), and Sharma et al. (74), used Minimum Norm Quadratic Unbiased Estimators (MINQUE) and Minimum Variance Quadratic Unbiased Estimators (MIVQUEO)

techniques, respectively, to estimate SOF variance components. Both obtained small estimates for the percentage of the total variance associated with SOF (1 to 4%) in milk production of Holstein cows. These estimates were smaller than those obtained with Henderson Method 1 on the same data sets.

Breed of SOF effect in approximately 12,000 first lactation records of Holstein heifers was investigated by Van Vleck (89) in unpublished research. The dependent variables were again three milk yield traits and two reproductive intervals. The fixed effects model used included HYS, group of SOC, size of cow and sex of calf in addition to breed of SOF. The SOF were Holstein active, Holstein young unproven, Jersey, Angus, and Hereford sires. Analysis of variance showed no significant differences among breed of SOF for milk yield traits, but significant differences at 5% level for gestation length and days open were found.

SOF and concurrent sire effects on milk yield, component yields and percentages, and days open were studied by Touchberry and Balaine (81) and Balaine and Touchberry (6, 7) in 1980. Results indicated that SOF had significant effects only on % fat in the first lactation, and on milk, fat and SNF yields and % protein in the second lactation. Effects on protein yield and days open were not significant in either lactation. Effects associated with

concurrent sire were significant for fat and protein percent in the first lactation only. These results agreed closely with those found by Quesnel (57).

Beef Breed Sires

Sire of fetus effects upon performance of the dam's current calf in beef cattle were studied by Shannon (73). Weight and condition score, adjusted for age, sex and sire of calf, and age of dam, were used as dependent variables. Results obtained indicated that the developing fetus affected performance of the nursing calf and this effect increased with fetal age. It appeared sires that produced desirable calves had positive effects on the performance of suckling calves. These findings, though not completely comparable, thus disagreed with results of Taylor et al. (79) with dairy cattle.

Effects of beef-sired fetuses on current and subsequent lactations of their dams were examined by Lindstrom et al. (44) using 16,572 milk records in Finland. Differences in 305-day milk yields between two consecutive lactations of Ayrshire cows sired by Ayrshire or beef bulls in the same herd were analyzed. Sire effects were studied by least squares procedures and variance component analyses in mixed models were performed using Henderson Method 3 (30).

When both dairy and beef bulls were included, the between sire variation accounted for 1.7 to 1.8% of the total variation in the milk yield of the dams in the concurrent lactation. The variation among beef bulls seemed to be somewhat smaller. The fact that variation among Ayrshire sires was larger for only untested than for progeny tested bulls indicated that nonrandom use of sires influenced comparisons of this kind. Lindstrom et al. (44) found evidence that beef sires had a detrimental influence on milk production of the cow in the concurrent lactation, but only for cows producing at higher levels or for younger parity animals. This influence disappeared in the subsequent lactation; i.e., beef SOF did not have a detrimental influence upon subsequent milk production of Ayrshire cows. These results agreed with those obtained by Van Vleck (89).

Summary

Based upon existing research on SOF effects on dairy cow performance, the following can be concluded:

1. When Henderson Method 1 was used as the variance component estimation technique, the estimates for SOF obtained by different authors, expressed as percentages of the total variance, differed widely (1 to 14%), but often were moderately large (1, 2, 5, 36, 37, 75, 76).

2. When Henderson Method 3 or equivalent mixed model methods were used, estimates obtained decreased considerably in range and magnitude (0 to 2%). The reduction probably was due to the ability of Method 3 to reduce correlated effects arising from nonrandom use of SOF across HYS and SOC (29, 37, 38, 44, 57, 58, 74, 92).
3. The effect of beef SOF upon milk production of dairy cows is not clear (42, 43, 44, 89).

Future research is needed to obtain unbiased variance component estimates for SOF effects. The use of Henderson Method 3, and MIVQUEO with comparable models, seems the approach indicated to obtain estimates free of confounded effects arising from covariances among other main effects.

Environmental Correlations

The influence that common conditions of environment, feeding and systems of management could have upon sets of half sisters biasing the estimates of the breeding value of their common sires was perhaps first discussed by Lush (45, 46) over 40 years ago. This was identified as a common environment path coefficient, and symbolized e^2. Lush (45) quoted unpublished research from Wright and other authors who found values of 0.15 to 0.30 for e^2. The fact that e^2 is a description of heterogeneity of conditions under which different sets of half sisters are kept was emphasized;

Lush explained that e^2 was the correction factor to adjust for heterogeneous environmental conditions among sets of half sisters when it was not possible to control them physically, as is the case in progeny testing.

Bereskin and Freeman (8) in 1965 recognized the importance of identifying sources of extraneous variability which bias estimates of sire breed values and reducing their effects where feasible. HYS was identified as the main source of bias. Using deviation records as a solution to the problem of residual variances, as proposed by Henderson (31), was criticized by Bereskin and Freeman (8), who considered this an incomplete solution. In practice, deviated records may not realize their full theoretical promise in removing HYS biases from production records (8, 48).

Bereskin and Lush (9, 10) reintroduced the importance of adjusting for residual variances that contribute to extraneous correlations among the daughters of the sires being tested. This extra correlation, which paternal sisters may have in common but which may be different from one set of paternal sisters to another, was symbolized then by c^2, and it was caused by factors other than sire alone. Correlated environmental effects, correlation between the breeding values of the mates of the sire, correlations between the breeding values of the sire and his mates, correlations involving both environmental and

genetic effects, distribution of the daughters among HYS groups, and the different adjustments used in computing the deviation records were identified by Bereskin and Lush (9, 10), as prominent possible causes which would contribute to c^2. The importance of including a nonzero value for c^2 when predicting sire breeding values was emphasized by Bereskin and Lush (9, 10), as they showed that as the size of c^2 increased, benefits of further increasing the number of daughters in a proof became increasingly smaller.

The magnitude of environmental correlations and their role in evaluation of sires used under conditions of AI in New York state were studied by Van Vleck (82, 83). Herdmate deviation methods of first lactation Holstein cows were sorted according to birth date of the cow, constructing different groups with different time intervals between them. Product moment correlations between daughter averages of the various pairs of groups were computed and these estimates were compared with expected correlations. Van Vleck (83) noticed that, under certain assumptions, the expression for actual correlations could be written as $s/(s + c + e/n)$, where s is the genetic variance, c is the environmental variance between paternal half-sibs, e is the residual variance, and if the estimate of s contained c, then the expected correlation based on those estimates would be $(s + c)/(s + c + e/n)$. Thus correlations actually computed should be $c/(s + c + e/n)$ smaller than expected

from the biased estimate of s, which really estimated s + c. Differences between actual and expected correlations were found to be practically zero. Therefore, Van Vleck (83) concluded that environmental correlations were small or nil for this population.

In another study Van Vleck (84) compared half-sib correlations in the same or different herds, and this difference was considered to estimate the environmental correlation, c^2, between paternal half-sibs in the same herd. Values found for these half-sib correlations were 0.168 for the same herd, which suggested that environmental correlation among paternal half-sibs in the same herd represented about 0.08 of the total variance.

The formula used presently by USDA (13, 14, 50, 51, 55) to compute the predicted differences is

$$\frac{\dfrac{Nh^2}{4 + (N-1)h^2 + \dfrac{4\Sigma n_i(n_i - 1)c^2}{N}}}{\text{Repeatability}} (D - AHMA) + .1\,(AHMA - E)$$

Predicted Difference

where

N = total number of progeny,

n_i = number of progeny in the i th herd,

h^2 = heritability from paternal half-sib estimates,

c^2 = environmental correlations among half-sibs,

D = daughters average,

AHMA = adjusted herdmates average

B = breed average.

Predicted Difference estimates expected performance of daughters of a sire when he is used in the general population. Repeatability measures accuracy of the correlation of a sire's daughters future performance and his estimated breeding value.

In both Predicted Difference as a whole, and Repeatability, c^2 plays an important role by adjusting for nonrandom components of environmental variance. In addition to strictly environmental similarities, c^2 includes everything else in the correlation between half-sisters besides the additively genetic part. The value of c^2, based on USDA research with nationwide samples of data, is 0.14 (55). The authors apparently did not describe exact methods used in this estimation; this value currently is used by USDA.

Environmental correlations from 158, 336 records were evaluated by Thomson and Freeman (80) in 1970. All records were standardized to 305-day, ME, and deviated from their adjusted herdmate average.

The random model used in this analysis was the following:

$$Y_{ijk\ell} = \mu + a_i + h_j + s_k + e_{ijk\ell} , \qquad [2.7]$$

where

$Y_{ijk\ell}$ = lactation records,

μ = population mean,

a_i = effect of the i th year-season,

h_j = effect of the j th herd,

s_k = effect of the k th sire,

$e_{ijk\ell}$ = residual error effect.

The records of any two paternal half-sisters would always fall in one of the classifications indicated in Table 2.1, where c^2 represented the additional covariance among half-sisters for being in the same herd.

Using Henderson Method 1 (30) variance components were estimated for model [2.7] and the following intraclass correlations were defined by Thomson and Freeman (80).

$$s = (\sigma_a^2 + \sigma_h^2)/(\sigma_a^2 + \sigma_h^2 + \sigma_s^2 + \sigma_e^2),$$

$$t = \sigma_h^2/(\sigma_a^2 + \sigma_h^2 + \sigma_s^2 + \sigma_e^2), \text{ and}$$

$$w = \sigma_a^2/(\sigma_a^2 + \sigma_h^2 + \sigma_a^2 + \sigma_e^2).$$

All available records were included in the analyses for the correlation s, but only one record was selected at random from each HYS and included in the analyses for the correlations t and w.

TABLE 2.1

Components of Variance for Paternal Half-Sibs

Herd	Year-Season	Components of Variance[1]
Same	Same	$\sigma^2_s + \sigma^2_a + \sigma^2_h + c^2\sigma^2_e$
Same	Different	$\sigma^2_s + \sigma^2_h + c^2\sigma^2_e$
Different	Same	$\sigma^2_s + \sigma^2_a$
Different	Different	σ^2_s

[1]From Thomson and Freeman (80)

Thomson and Freeman (80) noted that the correlation between two half-sisters in different herds was expected to be $1/4\ h^2$, whereas the correlation between two half sisters in the same herd also included t, a correlation from being in the same herd, and an additional correlation c^2 from environmental conditions common to the pair. From this the following intraclass correlations could be derived:

$$t_1 = 1/4\ h^2$$

$$t_2 = 1/4\ h^2 + t + c^2$$

and, c^2 an estimate of common correlation from common environment could be obtained as

$$c^2 = t_2 - t_1 - t.$$

Following this method Thomson and Freeman (80) obtained estimates of correlation from common environment over and above the common herd effect of 0.005 and 0.102 for ME and deviated records.

Environmental correlations among paternal half-sisters in the same herd were estimated by four methods by Arora and Freeman (4) in 1971. These authors recognized that the definition of c^2 has not been consistent and that these four methods did not estimate exactly the same parameter. They outlined the methods as follows:

1. Difference in intraclass correlation among daughters in one herd versus daughters in many herds. The method was introduced by Van Vleck (84). The value of c^2 is maximum when all daughters of a sire are in one herd, and it is zero when each daughter is in a different herd. Two data sets are created, one with all daughters of a sire in one herd, the other with each daughter in a different herd. The intraclass correlations for both data sets are estimated and the difference between these two is an estimate of c^2.

2. Intraclass correlation due to herds, year-season, and HYS. This method was introduced by Thomson and Freeman (80) and has been described thoroughly above.
3. Slight variation of (2) above. The intrasire correlation estimated from the single-herd sire analyses of variance (i.e., $\sigma_s^2/(\sigma_s^2 + \sigma_h^2 + \sigma_e^2)$) contains 1/4 h^2 and c^2. The intraclass correlation from half-sisters in different herds (i.e., $\sigma_s^2/(\sigma_s^2 + \sigma_e^2)$) is expected to be 1/4 h^2. The difference between these two correlations is an estimate of c^2.
4. Intraherd and intrasire correlation. The intraherd correlation within sires should be an estimator of c^2 among paternal sisters in the same herd. From an analysis of variance between sires, between herds nested in sires, and error term, the intraherd correlation can be computed as $\sigma_{h(s)}^2/(\sigma_s^2 + \sigma_{h(s)}^2 + \sigma_e^2)$.

Results obtained by Arora and Freeman (4) following these four methods are summarized in Table 2.2.

Factors that make sire genetic estimates representative of their genetic transmitting ability were reviewed by Norman (50) and Norman et al. (52) and the role that environmental correlations play on this was stressed. They noticed that if paternal half-sisters were

TABLE 2.2

Environmental Correlations (c^2) Between Paternal Half-Sisters

Trait[1]	Method 1	Method 2	Method 3	Method 4
ME records				
Milk	0.014	0.064	0.059	
Milk fat	0.046	0.095	0.068	
Deviation records				
Milk	0.016	-0.010	0.016	0.103
Milk fat	0.036	0.002	0.024	0.113

[1] From Arora and Freeman (4)

managed and fed more nearly alike than other cows in the same herd, their responses should be confounded with a herd x sire interaction component.

Theoretical connotations related with treating environmental correlation c^2 as an interaction between environment and SOC were extensively studied by Henderson (32, 33). As reported by Norman (50), most analyses that estimated variance components using Henderson Method 1 (30) found only small values for herd x sire interaction variance. Values were positive near zero or slightly negative for milk and fat yield and % fat, for Guernsey, Jersey and Holstein cows according to Legates and Verlinden (40), and Burdick and McGilliard (11). Using unadjusted records for milk and fat yield, herd x sire variance

components ranged from -0.9 to 2.9% of the total variance in work conducted by Allaire and Gaunt (3), even though in the case of adjusted records the estimates could not be considered small. Comparing variance component estimates for milk and fat yield using ME and herdmate deviation records, similar results were found by Norman et al. (53). Values were small and negative (-1 to -5%) for ME but large and positive (8 to 13%) for deviation records. These authors concluded that expressing records as deviations from herdmates may influence estimates of herd x sire component of variance.

Environmental correlations among progeny for type traits were estimated by Yao et al. (93) in 1979. Using Henderson Method 1 (30) and three of the approaches described by Arora and Freeman (4), higher values than those reported by them for milk and fat yields were found. The three methods estimated similar environmental correlations for final score (0.16 to 0.18), slightly lower for scorecard traits and usually less than 0.10 for descriptive traits. The authors concluded that for all type traits environmental correlations were too large to be ignored in sire evaluation procedures.

Using a slight variation of the method introduced by Thomson and Freeman (80), environmental correlations were estimated as difference between two intraclass correlations by Mbah (49) and Hargrove et al. (23) in 1981. The

original data set consisted of deviation records from first lactation Holstein cows in Pennsylvania. The mathematical model for each deviation record included only sire and cow, and Henderson Method 1 (30) was used to estimate variance and covariance components. The environmental correlations found were 0.08 to 0.11 for milk, fat and protein yields and -0.04 and -0.03 for % fat and protein. The negative estimates for the percentage traits were not explained.

Research conducted by Lee (39) is the only known attempt to use Henderson Method 3 (30) to estimate variance components for herd x SOC interaction. This research showed that Method 3 consistently yielded more positive or less negative estimates of interaction variances for MILKYD than Method 1.

As recognized by Arora and Freeman (4) neither the definition of c^2 or the approach followed to estimate it has been unique through the different studies on this subject. Most researchers (4, 23, 55, 80, 84) following different methods found small values for milk yield c^2 (0.00 to 0.10). Nevertheless, in the equation used to estimate sire breeding values for milk production, USDA includes c^2 with a value of 0.14 (55). If c^2 coefficient were truly smaller than 0.14, the value presently used, then values for predicted difference and repeatability that quantify breeding value of dairy sires would be less conservative. Therefore, the need of future research

directed to estimate values for c^2 accurately is evident. Despite the important ability of Henderson Method 3 (30) to obtain more nearly unbiased estimates for variance components, this method has not been used to estimate values of c^2, except for Lee (39).

Statistical Aspects of Variance Component Estimation

Estimation of a sire's genetic value is one of the most difficult problems in animal genetics. The fundamental problem is that true genetic values are unknown, and consequently estimates cannot be compared with true values to assess their reliability (34). Different methods have been proposed for sire evaluation (69). In the U.S. two methods, USDA-DHIA Modified Contemporary Comparison Methodology at USDA (Beltsville) and Linear Mixed Model Methodology at Cornell University (16, 47, 50), are widely accepted. Both methods try to separate environmental (HYS) from additive genetic (SOC) effects; the estimation of variance component for random effects in these mixed models is of major importance. It is in this context that the theory of variance components estimation in mixed models played an important role in development of sire evaluation methods. The contributions of dairy cattle geneticists to statistical theory improved estimation of variance components from highly unbalanced data sets, as is

the case for sire evaluation. Prominent Dairy Science researchers with significant contributions to variance component estimation are Henderson, Searle, and Harvey, among others.

ANOVA Approach

The theory of variance component analysis for balanced data sets was completed by Eisenhart (15). But it was Henderson (30) in 1953 who introduced the methods for estimation of variance and covariance components for data sets with unequal subclass numbers. Henderson (30) in a classic paper described the three subsequently well-known methods for estimating variance components in the non-orthogonal case, and he characterized the problems that can arise with use of these three methods and how to overcome them.

The three Henderson methods were reviewed and reformulated in matrix notation by Searle in 1968 (70). He introduced a fourth method, as a variant of Method 2, involving somewhat simpler calculations. A description of the three Henderson methods for estimating variance components follows (28, 30, 70, 71):

1. Henderson Method 1. Quadratic functions, analogous to either corrected or uncorrected sums of squares, are computed and equated to their expectations, obtained under the assumption that all effects are

random, and solved for the unknown variances. The numeric values of the coefficients in the formulas for expected mean squares usually can be obtained using the synthesis method developed by Hartley (24) and Rao (63). Method 1 cannot be used for mixed models; it is appropriate only for completely random models. If fixed effects exist, such data could be analyzed under Method 1 either by ignoring the fixed effects altogether or considering them random. The variance component estimators obtained in this way would be biased in both cases. The main advantage of this method is ease of comparison.

2. Henderson Method 2. This method was designed to overcome limitations of Method 1 for mixed models. First fixed effects are estimated, then the data are adjusted for these estimators, and finally variance components are obtained by Method 1. The resulting variance component estimators are not biased by the presence of fixed effects. The fact that this method cannot be used in the presence of interactions between fixed and random effects is perhaps the most important deficiency of Method 2.

3. Henderson Method 3. It is based on the reduction in sums of squares due to fitting constants. The

difference in the reduction sums of squares between a full model and a submodel judiciously chosen yields unbiased estimators of the variance components of the full model. In the case of mixed models these estimators are unbiased by any fixed effects that are included in the model. In the case of random models, even if the model is such that some fixed effects are correlated with some random effects, the estimators obtained for the variance components are unaffected by these correlations. A difficulty that arises in the use of Method 3 is that it can yield more equations than there are components to be estimated. Often, the use of Method 3 involves calculating generalized inverses of large matrices, in calculating not only the reduction in sums of squares but also the coefficients for the variance components. This can be done either using the synthesis method already mentioned (24, 63) or the unified procedure derived from the Abbreviated Doolittle and Square Root Methods proposed by Gaylor et al. (20). Goodnight and Speed (22) and Speed and Hocking (77) developed a highly efficient computational algorithm for Method 3 making it more suitable for use with large matrices.

Essentially, an analysis of variance technique without making any assumptions about the distribution of the variables has been used in the classical approach for estimation of variance components in mixed models and unbalanced data sets (18, 60, 62). Another method of estimation based on maximum likelihood estimation and with the assumption of normality of the variables was proposed by Hartley and Rao (25) to overcome the unknown theoretical properties of the former methods. Unfortunately, in this last method the computations involved are very complicated and computing cost can be high (60, 62, 72).

MINQUE and MIVQUE Approach

Extending the ideas of a previous article about heteroscedastic variances (59), an entirely new method called Minimum Norm Quadratic Unbiased Estimation (MINQUE) was proposed by Rao (60, 62) that under some assumptions could be used to estimate variance components. If MINQUE estimates are scaled by appropriate weights and with normal distribution assumptions, they provide Minimum Variance Quadratic Unbiased Estimators (MIVQUE) (61). The mathematical procedures to obtain MIVQUE were developed by Rao and others (56, 59, 60, 61, 62).

MIVQUE (with 0 priors), or simply MIVQUE0, uses the value 0 as prior value of the variance component to be estimated, except for the error term; therefore, it is

suitable for variance component estimation when those are expected to be small.

Hartley et al. (26) suggested the use of MINQUE proposed by Rao (60, 62) as an efficient approach for estimation of components of variance in large experimental or sample survey designs with a large number of random effects and/or random levels. Hartley et al. (26) proved that MIVQUEO are unbiased, locally best, admissible, and asymptotically consistent estimators.

Goodnight (21) developed a computational algorithm to obtain these Hartley MIVQUE estimators based on a multipassing technique that is able to calculate generalized inverses of large matrices that exceed the computer storage capacity.

Computing simulation was used by Lin and McAllister (41) to compare Henderson Method 3, maximum likelihood (ML), restricted maximum likelihood (REML) and MINQUE methods of variance components estimation in a mixed model with herd and sire. Method 3, REML and MINQUE estimates of error variance were consistently closer to the true parameters than corresponding estimators by ML. REML and MINQUE estimates of h^2 agreed more closely with the true parameter in most of the cases. The authors concluded that REML and MINQUE are probably superior to Method 3 in estimating variance components with mixed models because in Method 3 the sums of squares are obtained under a fixed

model whereas the expected sums of squares are derived under a mixed model. The authors concluded that MINQUE should be preferred over REML because of savings in computing time.

CHAPTER III
SIRE OF FETUS EFFECTS

Introduction

As early as 1915 Gaines (19) hypothesized that SOF, through the fetus, could affect milk yield of the dam. Nevertheless, Skjervold and Fimland (76) in 1975 published the first paper supported by statistical analysis of 69,501 first lactation records from Norway. Skjervold and Fimland expressed the SOF effect on the milk production of the dam as the coefficient of correlation between the phenotype of milk production of the dam and the genotype of SOF. Using Henderson Method 1 (30) the estimates of these correlations where approximately 0.1, which would be equivalent to assessing that SOF effect accounted for 0.01 of the total variance. Genetic correlation between breeding values of SOC and SOF was found to be zero and nonsignificant.

Adkinson et al. (2) published the first estimates of SOF effects from a large data set from the U.S.A. Using Henderson Method 1 also, they found that SOF accounted for approximately 10% of the total variance for ME MILKYD and FATYD and approximately 2% for DAYSOP. The variance component of the interaction SOC x SOF was zero for yields but accounted for almost 5% of the total variation for DAYSOP.

In a designed experiment a negative relationship between cows' production and genetic value of SOF for milk yield was found by Taylor et al. (79).

Quesnel et al. (58) were the first researchers to report Henderson Method 3 (30) variance component estimates for SOF. Unfortunately, the data set that they used was relatively small and the values obtained for production traits seemed to be suspiciously high; values for reproduction traits were smaller.

Henderson Method 1 and Method 3 variance components for three measures of milk yield and two reproductive traits were studied by Johnson and Van Vleck (36, 37, 38). When Method 1 was used, SOF accounted for almost 5% of the total variance for MILKYD, but, when Method 3 was used, the variance composition due to SOF reduced to approximately 1% for MILKYD. This reduction was explained by the authors as the result of capabilities of Henderson Method 3 of eliminating associative effects of nonrandom use of SOF across HYS and SOC.

Results similar to those reported by Johnson and Van Vleck were reported by Hayes et al. (29) using Henderson Method 3 in a Canadian data set. Percentage of the total variance accounted for by SOF was less than or equal to 1%, either for milk and component yields or percentages.

Attempts were made by Wickham (92) and Sharma et al. (74) to use MINQUE (60) and MIVQUE0 (26) approaches to

estimate SOF variance component. Using models comparable to those used with Henderson Method 1, estimates obtained were smaller.

Table 3.1 summarizes the results obtained in the most relevant studies about this subject.

TABLE 3.1

Existing Research Concerning SOF Effects on Milk Production

Authors	Method	SOF effect
Skjervold and Fimland (76)	Method 1	1% total variance
Adkinson et al. (2)	Method 1	10% total variance
Taylor et al. (79)	Regress. Analysis	- 1/3 SOF breeding value
Quesnel et al. (58)	Method 3	16% total variance
Johnson and Van Vleck (36)	Method 1	5% total variance
Johnson and Van Vleck (37)	Method 3	1% total variance
Hayes et al. (29)	Method 3	1% total variance
Wickham (92)	MINQUE	1% total variance
Sharma et al. (74)	MIVQUEO[1]	5% total variance

[1] Random model

From the existing research on SOF effects it can be concluded that when Henderson Method 1 was used the estimates obtained were generally moderately large. However, most estimates decreased considerably when Henderson Method 3 was used. The reduction probably was due to the ability of Method 3 to reduce correlated effects arising from use of SOF across HYS and SOC. Variance component estimates obtained using MIVQUEO techniques were smaller than those obtained with Method 1 with comparable models.

The purpose of the present research was to determine accurately the magnitude of the SOF variance relative to total variance in production and reproduction traits of dairy cattle. Henderson Method 3 are mixed model MIVQUEO techniques were used on large data sets and adequate models. Estimates obtained for SOF effects were expected to be unbiased by the existence of possible correlated effects arising from interactions among SOF, HYS and SOC.

Material and Methods

Data

The original data set used in this research consisted of 2,241,025 lactation records computed by the DHI program at the Dairy Records Processing Center, North Carolina State University, Raleigh. These records originated from

cows in North Carolina, South Carolina, Georgia and Florida between 1965 and 1982.

Different formats were used when the data were recorded on tapes. Therefore, considerable effort was necessary to modify all records to a unique format deleting all variables not needed for analyses to be performed.

Data were edited in different steps. First all records that were abnormal, three-times-a-day milking, nondairy breeds for SOC and SOF, or which contained missing information for milk production were eliminated. The data set obtained after this editing consisted of 614,766 records and called Data Set 1. It is presented in Table 3.2.

Identification of the service sire (SOF) in DHI records is quite imperfect. Often the sire is not identified at all; different identification systems are used (registration number, stud sire code, name, etc.); even the location of the identification within the field assigned in the record is not consistent. Unique identification of each sire was important for purposes of this research. If only registration number were accepted as correct sire identification, a great amount of information would have been lost. To avoid this, a tape containing registration number, stud sire code, and registered name of all dairy bulls used in AI between 1965 and 1982 (12) was merged with the tapes containing Data Set 1. An algorithm

TABLE 3.2

Data Sets, Distributed by State and Breed

State	n	%	Breed	n	%
Data Set 1[1]					
North Carolina	287,619	47	Holstein	505,179	82
South Carolina	104,702	17	Jersey	107,257	17
Georgia	127,474	21	Guernsey	2,302	<1
Florida	94,971	15	Brown Swiss	18	<1
Total	614,766	100	Total	614,766	100
Data Set 2					
North Carolina	62,710	49	Holstein	111,297	86
South Carolina	22,885	18	Jersey	17,814	14
Georgia	22,490	17	Guernsey	270	<1
Florida	21,296	16	Brown Swiss	0	0
Total	129,381	100	Total	129,381	100
Data Set 3					
North Carolina	50,959	49	Holstein	90,641	86
South Carolina	18,958	18	Jersey	14,147	14
Georgia	17,701	17	Guernsey	270	<1
Florida	17,440	16	Brown Swiss	0	0
Total	105,058	100	Total	105,058	100

[1] See text for description of data sets

converted as many as possible of all different sire identifications available to a uniform registration number. As a result, the final number of records obtained with SOF identification was considerably higher than the amount that would have been obtained using only those with a correct initial SOF identification. Records with missing or invalid values either for date of birth, calving date, or breeding date were eliminated at the same time. The data set obtained after this editing was called Data Set 2 and is presented in Table 3.2.

Finally, all records either with missing GESTL information or values larger than 300 or shorter than 250 days were eliminated. Simultaneously, in order to introduce some improvement in the highly sparse nature of these data sets (68), records corresponding to HYS, SOC or SOF with less than four observations were eliminated. The data set obtained was called Data Set 3 and also is in Table 3.2.

The subset of Jersey records from Data Set 3 was used in most analyses performed, and it was denoted Data Set 4. This can be considered a truly screened data set, such as those of Adkinson et al. (2) and Johnson and Van Vleck (37). Characteristics of this data set can be seen in Table 3.3. When computing limitations due to method of analysis did not allow use of the entire data set, it was subdivided by states, and analyses were performed on data

from each state separately. This partitioning of the Jersey data set can be seen in Table 3.4.

For data management purposes SAS82 (65, 66) computer software system and Utilities (54) package developed at University of Florida were used.

TABLE 3.3

Jersey Data Set, Means and Standard Deviations

Variable	Mean	s.e. of the mean	n
MILKYD[1]	4,390	10.5	14,147
FATYD[1]	207	0.62	
FAT%	4.72	0.01	
DAYSOP	117.6	0.50	
GESTL	279.6	0.05	
HYS	-	-	818
SOC	-	-	544
SOF	-	-	318

[1] in kg.

Analyses

The main goal of this research was to estimate variance components in random and mixed models. Henderson Method 1 and Method 3 (30) and MIVQUE0 (26) were used as variance component estimation procedures.

TABLE 3.4

Jersey Data Set by State

State	n	%
North Carolina	1,984	14
South Carolina	5,237	37
Georgia	2,677	19
Florida	4,249	30
Total	14,147	100

Model. Because the computing approach used did not allow for estimating variance components in models containing continuous effects, the response variables were adjusted first for these effects; then the residuals obtained were analyzed to estimate the variance components of the random discrete effects (64).

The general complete model for the approach as described above was

$$y = Xb + e,$$

where

y = n-vector of observed values: MILKYD, FATYD, FAT%, DAYSOP, and GESTL,

X = known n x p matrix,

b = p-vector of unknown parameters for the continuous fixed effects; it contains the linear, quadratic and cubic terms for age at parturition and length of record.

e = n-vector of random variables, $e \sim NID(0, \sigma^2 I)$, with $n = 5$ and $p = 6$.

Response variables were not adjusted for days pregnant due to the almost perfect correlation that existed between days pregnant and the dependent variable DAYSOP.

The residual vector from [3.1] was estimated

$$\hat{e} = y - X\hat{b} \qquad [3.2]$$

Finally, residual estimates obtained in [3.2] were used to estimate variance components. The general, completely random, full model was

$$\hat{e}_{ijk\ell} = \mu + A_i + B_j + C_k + AB_{ij} + AC_{ik} + BC_{jk} + ABC_{ijk} + e'_{ijk\ell} \quad , \qquad [3.3]$$

where

$\hat{e}_{ijk\ell}$ = residual estimate vector for MILKYD, FATYD, FAT%, DAYSOP, and GESTL as in [3.2],

μ = overall mean,

A_i = random effect of the i th HYS,

B_j = random effect of the j th SOC,

C_k = random effect of the k th SOF,

$e'_{ijk\ell}$ = random true residual,

AB_{ij}, AC_{ik}, BC_{jk} and ABC_{ijk} are the random interaction effects,

and the estimate

$$\hat{\underset{\sim}{\sigma}}^{2\prime} = \hat{\sigma}^2_A \; \hat{\sigma}^2_B \; \hat{\sigma}^2_C \; \hat{\sigma}^2_{AB} \; \hat{\sigma}^2_{AC} \; \hat{\sigma}^2_{BC} \; \hat{\sigma}^2_{ABC} \; \hat{\sigma}^2_{e'}$$

was obtained.

Henderson Method 1. To obtain Henderson Method 1 (30) estimates, the approach proposed by Harvey (27) was used. Following this approach all six possible combinations of nested effects were computed. The computed nested variance components were set equal to their expectations in terms of the cross-classified variances, and solving the set of six simultaneous equations the cross-classified variance components were obtained. Variance components for the nested models were obtained using PROC NESTED of SAS82 (67) computer software system, and solutions for the set of six simultaneous equations were obtained using PROC MATRIX, also from SAS82. These Henderson Method 1 estimates were obtained very economically, but, as indicated in Chapter II, if either some effect is not truly random, or some covariance between random effects exists, they would be biased.

Henderson Method 3. Since beginning this research it was clear that it would be prohibitively expensive to compute Henderson Method 3 (30) variance component estimates for all main effects, first and second order interactions of the complete model, given the large number of levels for main effects. Efforts were directed at obtaining Henderson Method 3 variance component estimates using submodels that would not overwhelm computing limitations. Therefore, during this first phase of research, estimation

of Henderson Method 3 variance components was attempted for models with only main effects.

PROC VARCOMP and PROC GLM from SAS82 (67) computer software system were used. The Sums of Squares Type I (Sequential Sums of Squares) provided by SAS82 for a completely random model with only main effects are indicated in Table 3.5.

TABLE 3.5

Type I SS and RSS for Main Effects in a Random Model

Effect	Type I SS	RSS
A	R(A/u)	R(u,A) - R(u)
B	R(B/u,A)	R(u,A,B) - R(u,A)
C	R(C/u,A,B)	R(u,A,B,C) - R(u,A,B)
e'	SSE	$y'y$ - R(u,A,B,C)

The Type I SS for each of the three main effects [3.3] and the respective Henderson Method 3 variance component estimates were obtained by running three analyses of variance; in each of them a different effect was fitted last after the other two effects were already in the model. The error term was the same in all three runs.

The Henderson Method 3 variance component estimates, as indicated in Chapter II, are unbiased either by presence of fixed effects in the model or when covariances between random effects exist. Unfortunately, they were more difficult to obtain since computer cost was high due to the large number of subclass levels.

MIVQUEO. As indicated in Chapter II, MIVQUEO (26) variance component estimates are appropriate when the parameter being estimated is expected to be small, as is the case for SOF variance components. Using a random model estimates obtained by MIVQUEO were comparable to estimates obtained with Henderson Method 1 (30). When MIVQUEO estimates were obtained in a mixed model, those estimates would be expected to be similar to estimates obtained with Henderson Method 3 (30).

PROC VARCOMP from SAS82 (67) computer software system was used to obtain MIVQUEO. For a completely random model only one run was necessary; given the computing efficiency of this approach, the use of a complete model with all main effects, first and second order interactions were possible. When fixed effects were present in the model, the computing cost increased notably. A run was needed for each effect, as with Henderson Method 3, and only estimates for the main effects were obtained.

Results and Discussion

The Henderson Method 1 (30) variance component estimates obtained using the complete model are shown in Table 3.6. It can be seen that some of the estimates were negative, interpretation of a negative estimate of a nonnegative parameter is obviously a problem (71). Even though a variance component is by definition positive, often estimates can be negative, especially when the data set is highly unbalanced, as it was in this case. In

TABLE 3.6

Henderson Method 1 Variance Component Estimates for Complete Model, Allowing Negative Estimates

Source[1,2]	MILKYD	FATYD	FAT%	DAYSOP	GESTL
A	31.2	61.3	78.6	5.9	2.1
B	21.1	32.6	38.1	1.4	2.5
C	11.8	23.2	24.3	0.5	9.7
AB	-16.2	-29.3	-34.1	0.1	3.6
AC	-7.1	-19.6	-25.2	-1.5	1.7
BC	-3.5	-3.8	3.2	3.4	2.6
ABC	0.3	5.8	-2.0	2.7	-4.0
e'	62.3	29.8	17.1	87.5	81.8
Total Variance	4556769	17683	.000117	3039	30.80

[1] Expressed as percentage of the total variance
[2] A = HYS, B = SCC, C = SOF and e' = Error

Table 3.7 negative estimates were set equal to zero. Although this seems a logical replacement, these estimates were no longer unbiased (71). In both cases the SOF component for production parameters were moderately large (9.3 to 24.3%) and for reproduction parameters were small or moderate (0.5 to 9.7%). Variance component estimates obtained for HYS and SOC were comparable to previous research (11, 90, 91).

TABLE 3.7

Henderson Method 1 Variance Component Estimates for Complete Model, Setting Negative Estimates Equal to Zero

Source[1,2]	MILKYD	FATYD	FAT%	DAYSOP	GESTL
A	24.6	40.1	48.7	5.8	2.0
B	16.6	21.4	23.6	1.4	2.4
C	9.3	15.2	15.0	0.5	9.3
AB	0	0	0	0.1	3.4
AC	0	0	0	0	1.7
BC	0	0	2.0	3.3	2.5
BC	0.2	3.8	0	2.7	0
e'	49.3	19.5	10.7	86.1	78.7
Total Variance	5779504	27003	.000188	3085	32.03

[1] Expressed as percentage of the total variance
[2] A = HYS, B = SOC, C = SOF and e' = Error

Results obtained for MILKYD and FATYD, when negative interaction component estimates were set equal to zero (9.3 to 15.2%), were very close to those obtained by Adkinson et al. (2) (8.2 to 13.9%), using the same model and restrictions. Johnson and Van Vleck (36, 37) obtained somewhat smaller estimates (2.7 to 4.8%) using a reduced model and not setting restrictions against negative estimates.

Close agreement existed between results obtained here for DAYSOP and GESTL (0.5 and 9.7%) and those obtained by Johnson and Van Vleck (-0.2 and 6.0% in (37), and 0.8 and 8.6% in (36)). Estimates obtained by Adkinson et al. (2) for DAYSOP were slightly higher (1.9 to 3.2%).

MIVQUEO (26) estimates for SOF components using a random complete model are shown in Table 3.8. In this case the Jersey data set was split by states due to computing limitations in performing the analyses with the entire set. Results are summarized using arithmetic means and means weighted by number of records in each state.

A random model was used in this analysis. Therefore, results are comparable to those obtained with Henderson Method 1 variance component estimation. Results obtained with MIVQUEO analysis for MILKYD, FAT%, DAYSOP and GESTL followed those obtained with Method 1 reasonably closely, but were smaller (7.4, 18.3, 0.0 and 13.0% vs. 11.8, 24.3, 0.5 and 9.7% for weighted means of MIVQUEO and Method 1).

TABLE 3.8

MIVQUEO for SOF, Random Complete Model

State[1]	MILKYD	FATYD	FAT%	DAYSOP	GESTL
North Carolina	2.7	5.5	9.5	-0.4	15.7
South Carolina	10.7	16.2	32.0	-0.2	10.3
Georgia	0.5	9.2	13.0	0.2	13.5
Florida	10.0	7.7	8.8	0.4	14.6
Arithmetic Mean	6.0	9.6	15.8	-0.0	13.5
Weighted Mean[2]	7.4	10.8	18.3	0.0	13.0

[1] Expressed as percentage of the total variance
[2] Weighting factor was the number of observations

However, MIVQUEO was definitely smaller for FATYD (10.8% vs. 23.2%).

These MIVQUEO's for MILKYD, FATYD and DAYSOP agreed closely with MIVQUEO estimates obtained by Sharma et al. (74) analyzing a Jersey data set (8.3, 12.2 and 0.03%) but are larger than those they obtained for MILKYD and FATYD analyzing a Holstein data set (4.5 and 4.9%).

Results obtained using Henderson Method 3 and MIVQUEO with a mixed model as variance component estimation procedures are shown in Tables 3.9 and 3.10. Again due to computing limitations, these estimates were obtained using a model containing only main effects and analyzing each

TABLE 3.9

Henderson Method 3 Variance Component Estimates for SOF, Model with Only Main Effects

State[1]	MILKYD	FATYD	FAT%	DAYSOP	GESTL
North Carolina	0.7	1.7	3.3	0.1	10.7
South Carolina	0.9	2.0	2.3	0.0	9.0
Georgia	3.0	3.9	5.6	-0.4	10.1
Florida	-0.6	1.3	2.0	2.4	12.0
Arithmetic Mean	0.9	2.2	3.3	0.5	10.5
Weighted Mean[2]	0.8	2.1	3.0	0.7	10.3

[1] Expressed as percentage of the total variance
[2] Weighting factor was the number of observations

TABLE 3.10

MIVQUEO for SOF, Mixed Model with Only Main Effects

State[1]	MILKYD	FATYD	FAT%	DAYSOP	GESTL
North Carolina	0.0	-0.1	0.0	0.1	15.4
South Carolina	0.7	1.0	0.0	0.2	11.4
Georgia	0.9	0.9	1.4	-0.2	16.3
Florida	1.0	1.6	2.4	0.5	11.3
Arithmetic Mean	0.7	0.9	1.0	0.2	13.6
Weighted Mean[2]	0.7	1.0	1.0	0.2	12.9

[1] Expressed as percentage of the total variance
[2] Weighting factor was the number of observations

state independently. As before, arithmetic and weighted means were used to summarize results.

It seems clear from this research that Henderson Method 1 estimates for production parameters were confounded with covariances existing between main effects, and were, therefore, inflated (9.3 to 24.3%). When either Henderson Method 3 estimates, or comparable MIVQUEO with mixed models, were obtained they were freed of any correlation that could exist among main effects, and the estimates obtained decreased considerably (0.7 to 3.3%). MIVQUEO with random models for production parameters (6.0 to 18.3%) fell in between Method 1 and Method 3 estimates.

Method 1, Method 3 and MIVQUEO estimates for reproduction parameters, either with random or mixed models, were similar. With all methods SOF effects were quite small for DAYSOP (0.0 to 0.7%), but appreciable for GESTL (9.3 to 13.6%). The considerable influence that SOF exerted upon GESTL is well known, i.e., parturition time to some degree is determined by the fetus and half of the fetal genotype comes from the sire.

It is tempting to compare results obtained in our research when Henderson Method 3 was ued with findings of Johnson and Van Vleck (37) using Method 3 also. Their results (0.8, -0.2 and 4.9% for ME MILKYD, DAYSOP and GESTL) were very similar to those found in this research. Nevertheless, in Johnson and Van Vleck's research, variance

components for SOF were estimated after fitting main effects and first order interaction for HYS and SOC

$$R(u,A,B,AB,C) - R(u,A,B,AB) = R(C/u,A,B,AB). \qquad [3.4]$$

Results obtained for SOF variance components using model [3.4] with Data Set 4 are shown in Table 3.11. Even though variability among states was high, results obtained for arithmetic and weighted means agreed with those shown in Table 3.9, being SOF variance components for MILKYD and GESTL slightly larger when model [3.4] was used.

TABLE 3.11

Henderson Method 3 Variance Component Estimates for SOF After Fitting HYS, SOC and HYS x SOC Interaction

State[1]	MILKYD	FATYD	FAT%	DAYSOP	GESTL
North Carolina	-5.0	-3.1	1.7	-3.1	13.2
South Carolina	1.5	1.2	-0.0	-0.0	10.0
Georgia	13.4	4.1	11.5	-0.0	12.0
Florida	-0.6	3.9	4.0	0.4	14.6
Arithmetic Mean	2.3	1.5	4.3	-0.7	12.5
Weighted Mean[2]	2.2	1.9	3.6	-0.3	12.2

[1] Expressed as percentage of the total variance
[2] Weighting factor was the number of observations

From comparing results shown in Tables 3.9 and 3.11 the effect of including HYS x SOC in the model to estimate SOF variance components was not clear. To elucidate this a model was used where variance components for SOF were estimated from the differences between reduction sums of squares of the complete model and a reduced model

$$R(u,A,B,AB,C,AC,BC,ABC) - R(u,A,B,AB)$$
$$= R(C,AC,BC,ABC/u,A,B,AB). \qquad [3.5]$$

With this approach, the variance component estimate for SOF included variance components of HYS x SOF, SOC x SOF and HYS x SOC x SOF. From [3.5] and Data Set 4, variance component estimates for SOF effects and the indicated interactions are in Table 3.12.

Even though considerable variability in Method 3 estimates of variance components by states existed for each variable, arithmetic and weighted means of SOF variances (with interactions) coincided closely with estimates when all four states were pooled together (Table 3.12) and were larger than estimates attributable to SOF alone (Table 3.9). These results seemed to indicate existence of real and measurable interactions between SOF effects and effects of HYS and SOC.

TABLE 3.12

Henderson Method 3 Variance Component Estimates for SOF Confounded with HYS x SOF, SOC x SOF and HYS x SOC x SOF Interactions

State[1]	MILKYD	FATYD	FAT%	DAYSOP	GESTL
North Carolina	-9.4	-3.5	11.3	19.0	15.3
South Carolina	0.1	9.6	-11.4	5.4	7.0
Georgia	21.9	5.9	36.2	17.2	7.7
Florida	7.5	21.7	9.8	4.9	22.2
Arithmetic Mean	5.0	8.4	11.5	11.6	13.0
Weighted Mean[2]	5.1	10.6	7.2	9.4	12.9
Pooled	4.1	13.9	6.5	5.6	16.9

[1] Expressed as percentage of the total variance
[2] Weighting factor was the number of observations

Summary and Conclusions

The original data set consisted of 2,241,025 lactation records collected by DHI programs in NC, SC, GA, and FL during 18 yr. Records that were abnormal, three-times-a-day milking, nondairy breeds for SOC and SOF, missing or invalid values for MILKYD, date of birth, calving date, or breeding date, missing GESTL, or values smaller than 250 or larger than 300 days, or with less than four counts for HYS, SOC and SOF were eliminated. SOF were uniquely identified with registration number. A subset of 14,147

Jersey records was obtained after editing. Henderson Method 1 and Method 3 (30) and MIVQUEO (26) techniques were used to estimate HYS, SOC and SOF variance components.

Using Henderson Method 1, SOF variance component estimates were moderately large (9 to 15%) for production parameters and GESTL and very small (0.5%) for DAYSOP. These results agreed closely with those reported by other authors. Variance component estimates attributable to SOF using Henderson Method 3 were considerably smaller (<1.0 to 3.0%) for production and reproduction traits. MIVQUEO estimates using mixed models were practically identical to those obtained with Method 3. MIVQUEO estimates using completely random models fell between Method 1 and Method 3 estimates. Reductions in production parameters estimates when Method 3 was used were attributed to the relative unbiasedness of Method 3 estimates due to covariances among HYS, SOC and SOF, i.e., SOF were not used at random among HYS and SOC.

Estimates obtained for SOF variances using a model where first order interactions between HYS and SOC were included agreed closely with those obtained with a model where only main effects were included. Estimates obtained for SOF which included first and second order interactions involving SOF and HYS and SOC, when Henderson Method 3 was used, were considerably higher than those for SOF alone. This suggested presence of real and measurable

interactions. Further research to determine the nature of these interactions was carried out in a subsequent research.

CHAPTER IV
ENVIRONMENTAL CORRELATIONS

Introduction

Over 40 years ago, Lush (45, 46) emphasized the importance of including a correction factor that describes the heterogeneity of the conditions under which different sets of half-sisters are kept in equations to estimate the breeding value of their common sires. He symbolized it by e^2 and quoted unpublished research from Wright and other authors where values of 0.15 to 0.30 were found for e^2.

The extra correlation which paternal half-sisters may have in common besides a common sire was symbolized as c^2 by Bereskin and Lush (9, 10) in 1965. Correlated environmental effects, correlation between the breeding values of the mates of the sire, correlations between the breeding values of the sire and his mates, correlations involving both environmental and genetic effects, distribution of the daughters among HYS groups, and the various adjustments used in computing deviation records were identified by Bereskin and Lush as prominent causes which increase c^2.

In the formula presently used by USDA to compute Predicted Difference and Repeatability, estimates of dairy

sires breeding value, c^2 of 0.14 is included (13, 14, 50, 51, 55). This value was obtained from a nationwide sample of data, but the authors apparently did not describe exact methods used in its estimation.

Environmental correlations among paternal half-sisters in the same herd were estimated by Van Vleck (84), Thomson and Freeman (80), Arora and Freeman (4) and Hargrove et al. (23). As recognized by Arora and Freeman, the definition of c^2 has not been consistent and at least five different methods to estimate c^2 can be identified. Each method estimates a somewhat different c^2. Four are variations about the central idea of obtaining c^2 as the difference between intraclass correlations among daughters in one herd versus daughters in many herds.

Factors that make sire genetic estimates representative of their genetic transmitting ability were reviewed by Norman (50), and the role that environmental correlations play in this was stressed. Norman identified failure to remove all HYS interaction effects, and paternal half-sisters being managed and fed more alike than other cows in the same HYS, as possibly the most important causes of environmental correlation. If paternal half-sisters are managed and fed more alike than other cows in the same HYS, their response should be confounded with HYS x SOC interactions.

Theoretical connotations of treating c^2 as an interaction between environment and SOC were studied extensively by Henderson (32, 33). Most analyses using Henderson Method 1 (30) to estimate variance components found only small values for herd x SOC interaction variance (3, 11, 40, 53). Research conducted by Lee (39) is the only known attempt to use Henderson Method 3 (30) to estimate variance components for herd x SOC interaction. This research showed that Method 3 consistently yielded more positive or less negative estimates of interaction variance for MILKYD than Method 1.

Table 4.1 summarizes results obtained by different researchers trying to estimate c^2 using different methods. Most of these values are smaller than 0.14 presently used by USDA to estimate dairy sires breeding values. If c^2 coefficients were truly smaller than 0.14, than values for Predicted Difference and Repeatability would be less conservative.

Falconer (17) mentioned complications that arise in estimating variance components if genotype-environment correlations and/or interactions exist, and how problems can be overcome by assigning these sources of variation to the genotype and environmental variance, respectively. He divided environmental variance into two components $V_E = V_{Ec} + V_{Ew}$. The common environment variance component

TABLE 4.1

Existing Research Concerning Environmental Correlation

Authors	Herd.Dev.Rec. MILKYD	Herd.Dev.Rec. FATYD	ME Records MILKYD	ME Records FATYD
Plowman and McDaniel (55)	0.140			
Van Vleck[1] (84)	0.082			
Thomson and Freeman[2] (80)	0.102		0.005	
Arora and Freeman[1] (4)	0.016	0.036	0.014	0.046
Arora and Freeman[2] (4)	-0.010	0.002	0.064	0.095
Arora and Freeman[3] (4)	0.016	0.024	0.059	0.068
Arora and Freeman[4] (4)	0.108	0.116		
Hargrove et al.[3] (23)	0.080	0.110		
Legates and Verlinden[5] (40)			0.010	0.005
Burdick and McGilliard[5] (11)			0.030	0.020
Allaire and Gaunt[5] (3)			0.030	
Norman et al.[5] (53)	0.08 to 0.15		-0.01 to -0.05	
Lee[5] (39)			0.002	
Lee[6] (39)			0.020	

[1] $c^2 = [(\hat{\sigma}^2_s - \hat{\sigma}^2_h)/\hat{\sigma}^2_t] - (\hat{\sigma}^2_s/\hat{\sigma}^2_t)$

[2] $c^2 = [h^2/4) + t + c^2] - [h^2/4] - t$

[3] $c^2 = [(h^2/4 + c^2] - (h^2/4)$

[4] $c^2 = \hat{\sigma}^2_{h(s)}/(\hat{\sigma}^2_s + \hat{\sigma}^2_{h(s)} + \hat{\sigma}^2_e$

[5] Method 1 variance component estimate for herd x SOC

[6] Method 3 variance component estimate for herd x SOC

V_{Ec} is caused by the common environment that members of a family share and cause similarity between members of the group. Within family sources of variation, V_{Ew} is independent of the relationship between members of the family.

If daughters of better sires are given preferential treatment, or a tendency exists to use better sires in better herds, i.e., SOC are not randomly used through HYS, this can be measured through the interaction HYS x SOC. On the other hand, quantification of interactions HYS x SOF and SOC x SOF would give useful information about existence of covariances between the main effects. These interactions may be pooled together and would constitute the environmental correlation c^2. Interactions of SOF with HYS and SOC and their role in the constitution of c^2 have never been studied.

In Chapter III, it was shown that when the SOF variance component is estimated by itself, it is considerably smaller than when combined with interactions. This suggests real existence of interactions among HYS, SOC and SOF.

For the present research, Henderson Method 3 and MIVQUEO with comparable models will be used to estimate variance components for HYS x SOC, HYS x SOF and SOC x SOF. The role of these interactions on the constitution of environmental correlation c^2 will be explored.

Material and Methods

Data

From an original data set consistng of 2,241,025 records produced by four major dairy breeds in the southeast, all records that were abnormal, three-times-a-day milking, nondairy breeds for SOC and SOF, missing or invalid values for MILKYD, date of birth, calving date, or breeding date, missing GESTL, or values shorter than 250 or longer than 300 days, and less than four observations for HYS, SOC and SOF were eliminated. SOF were uniquely identified with registration number. A subset of 14,147 Jersey records was obtained after screening. Editing procedures and characteristics of data sets were described in Chapter III.

Analyses

The principal purpose of this research was to estimate variance components for first order interactions of HYS, SOC and SOF using Henderson Method 3 (30) and MIVQUE0 (26) procedures and interpret these values in terms of environmental correlation c^2.

Analyses for variance components were performed on residuals obtained after adjusting the original lactation records for linear, quadratic and cubic terms of age at parturition and length of record. Details about models and computing techniques used were given in Chapter III.

From the beginning, it was clear that computing cost of obtaining variance components of the desired interactions would be very high. Nevertheless, the effort seemed worthwhile.

Henderson Method 3. To obtain Henderson Method 3 variance component estimates for first order interactions, PROC VARCOMP from SAS82 (67) computer software system with TYPE1 option (= Henderson Method 3) was used. For each interaction, a run was needed with a model where the first order interaction was fitted last after the three main effects were in the model. Type I Sums of Squares (Sequential Sums of Squares) provided by SAS82 for these models are indicated in Table 4.2. With this approach, $\hat{\sigma}^2_{AB}$, $\hat{\sigma}^2_{AC}$ and $\hat{\sigma}^2_{BC}$ were obtained.

The summation of variance components of main effects obtained in Chapter III, first order interactions obtained here and the residual in a model fitting main effects and first order interactions were considered as total variance.

MIVQUEO. The use of MIVQUEO (26), which considers a value of zero as a prior value of the variance components to be estimated, was justified given the small magnitude of the variance components expected for these interactions.

Variance components for first order interactions were calculated in a similar way that Method 3 estimated had been. PROC VARCOMP from SAS82 also was used, but in this case the MIVQUEO option was used. For each interaction, a

TABLE 4.2

Type I SS and RSS for a Random Model

Effect[1]	Type I SS	RSS
A	R(A/u)	R(u,A) - R(u)
B	R(B/u,A)	R(u,A,B) - R(u,A)
C	R(C/u,A,B)	R(u,A,B,C) - R(u,A,B)
AB	R(AB/u,A,B,C)	R(u,A,B,C,AB) - R(u,A,B,C)
	or	
AC	R(AC/u,A,B,C)	R(u,A,B,C,AC) - R(u,A,B,C)
	or	
BC	R(BC/u,A,B,C)	R(u,A,B,C,BC) - R(u,A,B,C)

[1] A = HYS, B = SOC and C = SOF

run was needed where all the other effects were considered fixed. Total variance components were estimated following the same approach used for Method 3 estimates.

Given the mixed nature of the models where MIVQUEO's were obtained, they should be comparable to variance component estimators obtained with Henderson Method 3.

Results and Discussion

Henderson Method 3 variance components estimates and MIVQUEO for first order interactions among HYS, SOC and SOF

and total variance by states are presented in Tables 4.3 and 4.4. Comparing these results with those obtained with Henderson Method 1 and MIVQUEO with comparable models obtained in Chapter III (Tables 3.6 and 3.8), it can be seen that values obtained here for first order interactions tended to be more positive and smaller in absolute value. These results agreed with findings reported by Lee (39) for the interaction herd x SOC.

Henderson Method 3 variance components for the interactions HYS x SOC, HYS x SOF and SOC x SOF expressed as percentage of the total variance by state are presented in Tables 4.5, 4.6 and 4.7. Results are summarized using arithmetic means of state percentages and weighted means, using the number of observations in each state as weighting coefficients.

All weighted means of first order interactions, for production as well as reproduction traits, were small and positive (0.4 to 6.5%). The small values obtained in this research for the interaction HYS x SOC in milk yield using Henderson Method 3 (1.1 to 4.1%) agreed with values for the interaction herd x SOC obtained by Lee (39) using Method 3, and other authors using Method 1 (3, 11, 39, 40, 53).

MIVQUEO for HYS x SOC, HYS x SOF and SOC x SOF are presented in Tables 4.8, 4.9 and 4.10. The weighted means for these variance components of first order interactions were similar to those obtained with Henderson Method 3 and

TABLE 4.3

Henderson Method 3 Variance Components Estimates for First Order Interactions

Effect	MILKYD	FATYD	FAT%	DAYSOP	GESTL
North Carolina					
HYS x SOC	-45414	-253.64	-.000003	92.95	0.4969
HYS x SOF	64471	16.35	.000001	66.52	2.3396
SOC x SOF	-72276	-393.15	.000000	164.88	0.5321
Total	2666909	6996	.000027	2672	28.12
South Carolina					
HYS x SOC	19667	218.83	.000001	88.78	1.0107
HYS x SOF	68685	477.28	.000001	-18.85	0.8607
SOC x SOF	-3684	372.91	.000001	132.03	1.6227
Total	3488569	8870	.000028	2931	29.53
Georgia					
HYS x SOC	23944	-275.12	.000003	-36.71	2.7602
HYS x SOF	594253	480.47	.000009	123.13	-0.4109
SOC x SOF	479893	179.61	.000006	-22.54	0.7539
Total	3762771	13915	.000109	2798	31.26
Florida					
HYS x SOC	120118	459.01	.000002	128.52	1.1069
HYS x SOF	313119	598.70	.000004	-65.66	1.3608
SOC x SOF	62409	525.28	.000003	223.29	0.0699
Total	3794561	8492	.000058	3314	32.52

TABLE 4.4

MIVQUEO for First Order Interactions

Effect	MILKYD	FATYD	FAT%	DAYSOP	GESTL
North Carolina					
HYS x SOC	-5097	-102.52	-.000002	77.00	0.2536
HYS x SOF	43178	28.52	.000001	37.10	1.9663
SOC x SOF	39666	-75.23	.000001	53.77	0.5283
Total	2522457	6310	.000026	2558	29.27
South Carolina					
HYS x SOC	-46858	38.63	-.000000	16.90	0.3023
HYS x SOF	66177	228.78	.000001	13.49	0.1956
SOC x SOF	19979	145.76	-.000000	-4.41	0.3212
Total	3175000	7518	.000023	2895	29.64
Georgia					
HYS x SOC	-2509	-258.84	.000001	-27.50	0.8476
HYS x SOF	33590	137.33	.000003	-21.07	-0.5795
SOC x SOF	152944	142.00	.000002	-25.09	0.4023
Total	3571941	10850	.000069	2713	32.44
Florida					
HYS x SOC	6250	45.19	.000000	42.04	0.7324
HYS x SOF	81223	166.59	.000001	-38.28	0.5181
SOC x SOF	-1030	-16.82	.000000	107.63	0.1900
Total	3755458	6676	.000041	3179	32.47

TABLE 4.5

Henderson Method 3 Variance Component Estimates for HYS x SOC

State[1]	MILKYD	FATYD	FAT%	DAYSOP	GESTL
North Carolina	-1.7	-3.6	-12.1	3.5	1.8
South Carolina	0.6	2.5	4.4	3.0	3.4
Georgia	0.6	-2.0	3.6	-1.3	8.8
Florida	3.2	5.4	4.9	3.9	3.1
Arithmetic Mean	0.7	0.6	0.2	2.3	4.3
Weighted Mean[2]	1.1	1.7	2.1	2.5	4.1

[1] Expressed as percentage of the total variance
[2] Weighting factor was the number of observations

TABLE 4.6

Henderson Method 3 Variance Component Estimates for HYS x SOF

State[1]	MILKYD	FATYD	FAT%	DAYSOP	GESTL
North Carolina	2.4	0.2	3.7	2.5	8.3
South Carolina	2.0	5.4	5.7	-0.6	2.9
Georgia	15.8	3.4	8.9	4.4	-1.3
Florida	8.2	7.1	7.1	-2.0	4.2
Arithmetic Mean	7.1	4.0	6.4	1.1	3.5
Weighted Mean[2]	6.5	4.8	6.4	0.4	3.3

[1] Expressed as percentage of the total variance
[2] Weighting factor was the number of observations

TABLE 4.7

Henderson Method 3 Variance Component Estimates for SOC x SOF

State[1]	MILKYD	FATYD	FAT%	DAYSOP	GESTL
North Carolina	-2.7	-5.6	1.1	6.2	1.9
South Carolina	-0.1	4.2	4.8	4.5	5.5
Georgia	12.8	1.3	5.9	-0.8	2.4
Florida	1.6	6.2	5.5	6.7	0.2
Arithmetic Mean	2.9	1.5	4.3	4.1	2.5
Weighted Mean[2]	2.5	2.9	4.7	4.4	2.8

[1] Expressed as percentage of the total variance
[2] Weighting factor was the number of observations

TABLE 4.8

MIVQUE0 for HYS x SOC

State[1]	MILKYD	FATYD	FAT%	DAYSOP	GESTL
North Carolina	-0.2	-1.6	-7.7	3.0	0.9
South Carolina	-1.5	0.5	-0.0	0.6	1.0
Georgia	-0.1	-2.4	1.4	-1.0	2.6
Florida	0.2	0.7	0.0	1.3	2.3
Arithmetic Mean	-0.4	-0.7	-1.6	1.0	1.7
Weighted Mean[2]	-0.5	-0.4	-0.8	0.8	1.7

[1] Expressed as percentage of the total variance
[2] Weighting factor was the number of observations

TABLE 4.9

MIVQUEO for HYS x SOF

State[1]	MILKYD	FATYD	FAT%	DAYSOP	GESTL
North Carolina	1.7	0.5	3.8	1.4	6.7
South Carolina	2.1	3.0	4.3	0.5	0.7
Georgia	0.9	1.3	4.3	-0.8	-1.8
Florida	2.2	2.5	2.4	-1.2	1.6
Arithmetic Mean	1.7	1.8	3.7	-0.0	1.8
Weighted Mean[2]	1.8	2.2	3.7	-0.1	1.3

[1] Expressed as percentage of the total variance
[2] Weighting factor was the number of observations

TABLE 4.10

MIVQUEO for SOC x SOF

State[1]	MILKYD	FATYD	FAT%	DAYSOP	GESTL
North Carolina	1.6	-1.2	3.8	2.1	1.8
South Carolina	0.6	1.9	0.0	-0.2	1.1
Georgia	4.3	1.3	2.9	0.9	1.2
Florida	-0.0	-0.3	0.0	3.4	0.6
Arithmetic Mean	1.6	0.4	1.7	1.6	1.2
Weighted Mean[2]	1.3	0.7	1.1	1.4	1.1

[1] Expressed as percentage of the total variance
[2] Weighting factor was the number of observations

followed the same trends, but they tended to be slightly smaller and some were negative (-0.8 to 3.7%).

Weighted means of Method 3 and MIVQUEO variance components for the interaction HYS x SOF (0.4 to 6.5% and -0.1 to 3.7%) and SOC x SOF (2.5 to 4.7% and 0.7 to 1.4%) tended to be slightly larger or more positive than those of HYS x SOC (1.1 to 4.1% and -0.8 to 1.7%) even though they were also small.

As discussed in Chapter III, Henderson Method 3 variance components estimated for SOF pooled together with HYS x SOF, SOC x SOF and HYS x SOC x SOF were larger than the variance components for SOF alone. This result suggested existence of sizable interactions among HYS, SOC and SOF. Differences between Method 3 variance components for SOF alone and SOF confounded with interactions, expressed as percentage of the total variance, were 4.3, 8.5, 4.2, 8.7 and 2.6% for MILKYD, FATYD, FAT%, DAYSOP and GESTL weighted means, respectively. Summing Method 3 variance components estimates for interactions HYS x SOF and SOC x SOF resulted in estimates of 9.0, 7.7, 11.1, 4.8 and 6.1%. Differences between these two sets of values may be attributed to the variance components of second order interaction HYS x SOC x SOF.

Summation of variance components for first order interactions among HYS, SOC and SOF expressed as percentage of the total variation, is a good estimate of environmental

correlation c^2. These interactions suggest presence of covariances, i.e., departure from randomness either in the manner in which SOC are distributed among HYS, SOF in HYS or SOF in SOC. Therefore, the summation of these interactions, with allowances for have being obtained each one in a different analysis, should be a good description of the common environmental correlation c^2. Values obtained for the summation of Method 3 variance components for the three first order interactions were 10.1, 9.4, 13.2, 7.3 and 10.2% for MILKYD, FATYD, FAT%, DAYSOP and GESTL weighted means, respectively. Summation of MIVQUEO values for the same effects were 2.6, 2.5, 4.0, 2.1 and 4.1%.

It seems quite possible that the value of 0.14 presently used by USDA to adjust for environmental correlation c^2 is biased slightly upwards. Values obtained in the present research with Henderson Method 3 and MIVQUEO, as well as values obtained by other authors (3, 11, 39, 40, 53) with Method 1, indicate that the value to consider for c^2 should be somewhat smaller. The implication that a smaller value for c^2 would have increasing values of Predicted Difference and Repeatability in sire testing is evident.

During the present research efforts were made to take into consideration the highly sparse nature of the data set. Nevertheless, future research oriented to obtaining

Henderson Method 3 estimates of c^2 in connected data sets and studies of effects of disconnectedness on the estimates would be of great importance.

Summary and Conclusions

A highly screened data set composed of 14,147 Jersey records were used in this research. Henderson Method 3 and MIVQUEO with mixed models techniques were used to estimate variance composition of first order interactions among HYS, SOC and SOF for MILKYD, FATYD, FAT%, DAYSOP and GESTL. All variance components were expressed as percentage of the total variance.

Small values obtained for the Method 3 variance components and MIVQUEO of the interaction HYS x SOC (1.1 to 4.1% and -0.8 to 1.7%) agreed with estimates of herd x SOC interaction obtained by Lee (39) using Method 3, and other authors using Method 1 (3, 11, 39, 40, 53). Variance components for HYS x SOF (0.4 to 6.5% and -0.1 to 3.7%) and SOC x SOF (2.5 to 4.7% and 0.7 to 1.4%) interactions were larger and more consistently positive. Method 3 and MIVQUEO estimates for the three interactions followed the same trend, but the MIVQUEO were of smaller scale.

Summation of the three first order interactions is a good estimate of the environmental correlation c^2. These interactions indicate existence of covariances between HYS and SOC, HYS and SOF, and SOC and SOF. Method 3 and

MIVQUEO estimates of c^2 obtained following this approach were 10.1, 9.4, 13.2, 7.3, 10.2% and 2.6, 2.5, 4.0, 2.1, 4.1% for MILKYD, FATYD, FAT%, DAYSOP, and GESTL.

Results obtained in this research using Henderson Method 3 and MIVQUEO with mixed models, in agreement with results obtained by other authors using Method 1, indicated that the value of 0.14 for c^2 presently used by USDA to evaluate sires is biased slightly upwards. With the use of smaller values of c^2 Predicted Difference and Repeatability coefficients would be less conservative. Further research on effects of disconnectedness in data sets upon estimation of c^2 seems appropriate.

CHAPTER V
COMPUTING EFFICIENCY OF HENDERSON METHOD 3 AND MIVQUEO

Introduction

Even though the Henderson Method 3 (30) variance components estimation procedure has many desirable properties, it has not been used extensively in the past for animal research due to the computing limitations found in calculating generalized inverses of large matrices (70). Recently, with availability of new computing facilities with greater memory and speed, Henderson Method 3 is being used more frequently.

MIVQUEO (26) is believed to be characterized by shorter computing time requirements than Method 3 to perform the same estimations (21, 26, 67). Quaas and Bolgiano (56) proved that the efficiency of estimation of MIVQUEO over Method 3 is better only when the variance being estimated is very small (<.01). Lin and McAllister (41) in simulation research showed that MINQUE and REML provided very similar estimators but MINQUE was preferable to REML because of savings in computing time. According to Lin and McAllister, both estimators MINQUE and REML were superior to Method 3. Goodnight (21) noticed that

computing time needed to obtain Method 3 and MIVQUE0 estimators was proportional to numbers of observations and number of levels of each effect, but he did not specify the coefficient of proportionality.

A comparison of needs of computing time when the same parameters are estimated with Method 3 and MIVQUE0 would be useful to help to determine what method of estimation to use in future research.

Material and Methods

Characteristics of data sets and models used in this research are described fully in Chapters III and IV. In each analysis the effects of interest, main effects or first order interactions, were fitted last and considered random, after other effects were already fitted and considered fixed. PROC VARCOMP from SAS82 (67) computer software system was used for variance estimation with Method 3 and MIVQUE0 methodology. Computations were performed on an IBM 3081D mainframe computer with Multiple Virtual Storage Extended Architecture (MVS/XA) operating system. A total of over 19 hours of CPU time was used in analyses reported in this research.

Results and Discussion

Computing time in CPU seconds used in estimation of variance components for main effects with Henderson

Method 3 and MIVQUEO are presented in Table 5.1. The region of memory used changed in each run, as is indicated in the table.

TABLE 5.1

Computing Time for Method 3 and MIVQUEO Estimating Variance Components of Main Effects

State[1]	Effect	Class Num.	D.F.	Memory K	CPU Seconds Method 3	CPU Seconds MIVQUEO
NC (1984)	HYS	164	160	1144	40.53	40.87
	SOC	198	193	1144	40.14	38.02
	SOF	111	104	1144	40.45	39.42
SC (5237)	HYS	257	253	1800	180.72	174.75
	SOC	334	332	1800	180.23	163.16
	SOF	194	189	1800	182.13	171.88
GA (2677)	HYS	211	208	1384	56.67	57.53
	SOC	188	187	1384	56.22	55.19
	SOF	133	130	1384	56.45	54.17
FL (4249)	HYS	186	183	1258	49.60	51.13
	SOC	203	201	1258	49.72	47.59
	SOF	113	109	1258	49.76	48.63
TOTAL					982.62	942.34

[1] Number of observations in parentheses.

It can be seen that number of levels of fixed and random effects has greater weight in determining computing time needed for the estimations than number of observations. The difference between Method 3 and MIVQUEO in total computing time used for the complete analysis was

very small (982.62 vs. 942.34 CPU seconds, difference 40.28 CPU seconds), slightly favoring MIVQUE0.

Use of computing time in estimation of variance components for interactions is presented in Table 5.2. All runs were performed in a region of 1800 K of computing memory.

TABLE 5.2

Computing Time for Method 3 and MIVQUE0 Estimating Variance Components of First Order Interactions

State[1]	Effect	Subclass Num.	Filed %	D.F.	CPU Seconds (1800K) Method 3	MIVQUE0
NC	HYSxSOC	1127	3.4	752	1292.03	879.12
(1984)	HYSxSOF	593	3.2	310	393.86	340.10
	SOCxSOF	960	4.3	637	925.69	686.12
SC	HYSxSOC	2576	3.0	1116	12085.29	6894.75
(5237)	HYSxSOF	1405	2.8	339	3341.34	2569.77
	SOCxSOF	2589	3.9	856	12151.07	6976.27
GA	HYSxSOC	1445	3.6	1026	2450.78	1557.65
(2577)	HYSxSOF	837	2.9	488	822.23	661.73
	SOCxSOF	1181	4.7	850	1592.54	1123.95
FL	HYSxSOC	1517	4.0	1116	2567.51	1571.00
(4249)	HYSxSOF	637	3.0	339	474.16	415.25
	SOCxSOF	1186	5.1	856	1510.78	1045.40
TOTAL					42058.06	24721.12

[1] Number of observations in parentheses.

The number of levels of fixed effects was the same in the estimation of variance components for the different

interactions within each state. Therefore, computing time was essentially determined by number of levels of the random effect, i.e., the interaction subclass numbers. The effect of number of observations on computing time is minimum. In this case, the difference between Method 3 and MIVQUE0 in total computing time used for the complete analysis was large (42,058.06 vs. 24,721.12 CPU seconds, difference 17,336.94 CPU seconds).

Comparing Tables 5.1 and 5.2, it can be seen that differences in computing time between Method 3 and MIVQUE0 increased geometrically with total time used (40.28 and 17,336.94 CPU seconds). It can, therefore, be concluded that efficiency of MIVQUE0 over Method 3 in use of computing time increases with the level of total computing time used; i.e., savings of computing time for the use of MIVQUE0 increased with size of analysis.

Summary and Conclusions

Use of CPU time for variance components estimation was compared with Henderson Method 3 and MIVQUE0 techniques using PROC VARCOMP of SAS82 computer software system in a IBM 3081D mainframe computer with MVS/XA. Results obtained indicated that MIVQUE0 was more efficient than Method 3 in use of computer time, and its efficiency increased with size of analysis.

CHAPTER VI
EFFECTS OF MISIDENTIFICATION OF SIRE OF FETUS ON ESTIMATES OF SIRE OF FETUS VARIANCE

Introduction

In 1968, Johansson and Rendel (35) reported that in Scandinavian countries approximately 4% of dairy records have incorrect identification for either sire or dam of the cow that produced the record. Obviously, the exact fraction of misidentification for sires and dams cannot be estimated from production records.

Van Vleck (85, 86) was able to identify four different models that described the problem of SOC misidentification. He reported that estimates of sire components of SOC and error were biased for all four models when SOC misidentification occurred. These biased estimates led to underestimation of heritability from intrasire correlation, and the degree of bias for heritability estimates was proportional to the fraction of SOC misidentified.

As concluded in Chapter III, even though SOF effects constituted only a small portion of the total variance, it would be interesting to know from a theoretical point of view how estimates of variance components are affected when some degree of misidentification exists for SOF, and the

consequences if any of this on heritability estimates for production and reproduction traits. Previous research about this problem is not known and it was the subject of the present chapter.

Material and Methods

Effects of SOF misidentification on estimates of SOF variance were studied using the following mathematical model

$$Y_{ijm} = \mu + B_i + C_j + e_{ijm} \qquad [6.1]$$

where

Y_{ijm} = response variable

μ = constant

B_i = fixed effect due to the i th SOC, i=1,2,...,s

C_j = random effect due to the j th SOF, j=1,2,...,r

e_{ijm} = random error effect, m=1,2,...,n_{ij}

and $C_j \sim NID(0, \sigma_c^2)$, $e_{ijm} \sim NID(0, \sigma_e^2)$, and $Cov(C_j, e_{ijm}) = 0$.

The assumptions were that B_i was properly identified every time, but some degree of misidentification existed for C_j. The problem of misidentification of C_j was classified into two general cases:

1. Within each SOF group when all misidentified fetuses were from the same sire, the model for the i th group was

$$Y_{ij\cdot} = n_{ij}\mu + n_{ij}B_i + n_{ij\ell}C_j + (n_{ij} - n_{ij\ell})C_{jk} + e_{ij\cdot} \quad [6.2]$$

where $n_{ij\ell}$ fetuses were properly identified as having the j th sire and $n_{ij} - n_{ij\ell}$ fetuses were being identified as having the j th sire but actually belonged to the jk th sire.

2. Within each SOF group when each misidentified fetus was from a different sire, the model for the ij th group was

$$Y_{ij\cdot} = n_{ij}\mu + n_{ij}B + n_{ij\ell}C_j + \sum_{k=n_{ij\ell}+1}^{n_{ij}} C_{jk} + e_{ij\cdot} \quad [6.3]$$

where $n_{ij\ell}$ fetuses were properly identified as having the j th SOF but $n_{ij} - n_{ij\ell}$ fetuses each had a different SOF than the j the SOF.

Henderson Method 1 was used in this research and the ANOVA table used is presented in Table 6.1.

Results and Discussion

Expectations of quadratic functions analogous to sums of squares are presented in Table 6.2.

To simplify the model, it was assumed that $n_{ij} = n$ and if the fraction of correctly identified SOF was p, then

TABLE 6.1

ANOVA Table to Study Effects of SOF Misidentification on Estimates of Variance Components

Source of Variation	SS
Total	$\sum_{ijm} y_{ijm}^2 - CF$
SOC	$\sum_{i} \frac{y_{i\cdot\cdot}^2}{n_{i\cdot}} - CF$
SOF	$\sum_{j} \frac{y_{\cdot j\cdot}^2}{n_{\cdot j}} - CF$
Residual	$\sum_{ijm} y_{ijm}^2 - \sum_{i} \frac{y_{i\cdot\cdot}^2}{n_{i\cdot}} - \sum_{j} \frac{y_{\cdot j\cdot}^2}{n_{\cdot j}} + CF$
Lack of fit	$\sum_{ij} \frac{y_{ij\cdot}^2}{n_{ij}} - \sum_{i} \frac{y_{i\cdot\cdot}^2}{n_{i\cdot}} - \sum_{j} \frac{y_{\cdot j\cdot}^2}{n_{\cdot j}} + CF$
Pure error	$\sum_{ijm} y_{ijm}^2 - \sum_{ij} \frac{y_{ij\cdot}^2}{n_{ij}}$

TABLE 6.2

Quadratic Functions for Models
with SOF Misidentifications

A. $E\left[\sum_{j} \frac{y_{\cdot j \cdot}^2}{n_{\cdot j}}\right]$

Case 1: $n_{\cdot\cdot}\mu^2+\sum_{ij} n_{ij}B_i^2+ \frac{\sum_{ij} n_{ij}^2+2\sum_{ij} n_{ij\ell}(n_{ij\ell}-n_{ij})}{n_{\cdot\cdot}} \sigma_c^2+r\sigma_e^2$

Case 2: $n_{\cdot\cdot}\mu^2+\sum_{ij} n_{ij}B_i^2+ \left[\frac{\sum_{ij} n_{ij}+\sum_{ij} n_{ij\ell}(n_{ij\ell}-1)}{n_{\cdot\cdot}}\right]\sigma_c^2+r\sigma_e^2$

B. $E(CF) = E\left[\frac{y_{\cdot\cdot\cdot}^2}{n_{\cdot\cdot}}\right]$

Case 1: $n_{\cdot\cdot}\mu^2+ \frac{\sum_{ij} n_{ij}^2+2\sum_{ij} n_{ij\ell}(n_{ij\ell}-n_{ij})}{n_{\cdot\cdot}} \sigma_c^2+\sigma_e^2$

Case 2: $n_{\cdot\cdot}\mu^2+ \frac{\sum_{ij} n_{ij\ell}^2+\sum_{ij} (n_{ij}-n_{ij\ell})}{n_{\cdot\cdot}} \sigma_c^2+\sigma_e^2$

C. $E\left[\sum_{ij} \frac{y_{ij\cdot}^2}{n_{ij}}\right]$

Case 1: $n_{\cdot\cdot}\mu^2+\sum_{ij} n_{ij}B_i^2+\left[n_{\cdot\cdot}+2\sum_{ij} \frac{n_{ij\ell}(n_{ij\ell}-n_{ij})}{n_{ij}}\right]\sigma_c^2+c\sigma_e^2$

Case 2: $n_{\cdot\cdot}\mu^2+\sum_{ij} n_{ij}B_i^2+\left[\sum_{ij} \frac{n_{ij\ell}(n_{ij\ell}-1)}{n_{ij}} c\right]\sigma_c^2+c\sigma_e^2$

TABLE 6.2-continued.

D. $E\left[\sum_{ijm} y_{ijm}^2\right]$	
Cases 1 and 2:	$n..\mu^2+\sum_{ij} n_{ij}B_i^2+n..\sigma_c^2+n..\sigma_e^2$

Note: Proofs of formulas for Case 1 are given in the Appendix. Proofs of formulas for Case 2 are extensions of these.

$n_{ij\ell} = pn$. With these assumptions, expectations obtained are shown in Table 6.3.

From formulas in Table 6.3, it can be seen that when SOF were misidentified (i.e., $p < 1$), the expectation of the error term variance component was biased upwards. The expectation of SOF variance component could not be found because the fixed effect SOC did not vanish in the respective quadratic forms. The inadequacy of Henderson Method 1 to estimate variance components in mixed models is well known.

TABLE 6.3

Expectations of SOF and Error Term Components of Variance for Models with SOF Misidentification

Case	$E[\hat{\sigma}^2_c]$	$E[\hat{\sigma}^2_e]$
1	It cannot be found because the fixed effect B does not vanish in the respective quadratic forms (Henderson Method 1)	$\frac{n(1-x)}{n-1}\sigma^2_c + \sigma^2_e$
2		$\frac{n-z}{n-1}\sigma^2_c + \sigma^2_e$

$z = p^2n - p + 1$

$x = 1 + 2p(p - 1)$

p = proportion of correctly identified SOF

$1 - p$ = proportion of incorrectly identified SOF

pn = total number of correctly identified SOF

Note: Proofs of formulas for expectations of error for Case 1 is given in the Appendix. Proofs of formulas for Case 2 is an extension of this.

Summary and Conclusions

Effects of SOF misidentification on estimates of SOF variance components using Henderson Method 1 was studied. It was found that as a result of SOF misidentification the error mean square was no longer an unbiased estimate of the error variance and it was biased upwards. However, expected values for SOF variance components could not be obtained due to the existence of the fixed effect SOC in the model.

CHAPTER VII
SUMMARY AND CONCLUSIONS

The original data set consisted of 2,241,025 lactation records collected by DHI program in NC, SC, GA, and FL during 18 years.

Records that were abnormal, three-times-a-day milking, nondairy breeds for SOC and SOF, missing or invalid values for MILKYD, date of birth, calving date, or breeding date, missing GESTL, or values shorter than 250 or larger than 300 days, and categories with less than four records for HYS, SOC and SOF were eliminated. SOF were uniquely identified with registration number. A subset of 14,147 Jersey records was obtained after editing. Henderson Method 1 and Method 3 (30) and MIVQUE0 (25) techniques were used to estimate HYS, SOC and SOF variance components.

Using Henderson Method 1, SOF variance component estimates were moderately large (9 to 15%) for production parameters and GESTL and very small (0.5%) for DAYSOP. These results agreed closely with those reported by other authors.

Variance component estimates attributable to SOF using Henderson Method 3 were considerably smaller (<1.0 to 3.0%) for production and reproduction traits. MIVQUE0 estimates

using a mixed model were practically identical to those obtained with Method 3. MIVQUEO estimates using completely random models fell between Method 1 and Method 3 estimates. The reduction in production variance component estimates when Method 3 was used was attributed to the relative unbiasedness of Method 3 estimates in the presence of possible covariances among HYS, SOC and SOF, i.e., SOF apparently were not used at random among HYS and SOC.

Estimates obtained for SOF variance composition using a model where the first order interaction between HYS and SOC was included agreed closely with those obtained with a model where only main effects were used. Estimates obtained for SOF which included first and second order interactions among SOF and HYS and SOC, when Henderson Method 3 was used, were considerably larger than those for SOF alone. This suggested existence of sizable interactions.

Henderson Method 3 and MIVQUEO with mixed model techniques were used to estimate variance composition of first order interactions among HYS, SOC and SOF for MILKYD, FATYD, FAT%, DAYSOP and GESTL. All variance components were expressed as percentages of the total variance. Small values obtained for the Method 3 variance components and MIVQUEO of the interaction HYS x SOC (1.1 to 4.1% and -0.8 to 1.7%) agreed with estimates of herd x SOC interaction obtained by Lee (39) using Method 3, and other authors

using Method 1 (3, 11, 39, 40, 53). Variance components for HYS x SOF (0.4 to 6.5% and -0.1 to 3.7%) and SOC x SOF (2.5 to 4.7% and 0.7 to 1.4%) were larger and more consistently positive. Method 3 and MIVQUE0 estimates for the three interactions followed the same trend, but the MIVQUE0 were of smaller scale.

Summation of the three first order interactions is a good estimate of the environmental correlation c^2. These interactions indicate existence of covariances between HYS and SOC, HYS and SOF, and SOC and SOF. Method 3 and MIVQUE0 estimates of c^2 obtained following this approach were 10.1, 9.4, 13.2, 7.3, 10.2% and 2.6, 2.5, 4.0, 2.1, 4.1% for MILKYD, FATYD, FAT%, DAYSOP, and GESTL.

Results obtained in this research using Henderson Method 3 and MIVQUE0 with mixed models, in agreement with results obtained by other authors using Method 1, indicated that the value of 0.14 for c^2 presently used by USDA to evaluate sires is biased slightly upwards. With the use of smaller values for c^2 Predicted Difference and Repeatability, coefficients would be less conservative.

Use of CPU time was compared for variance components estimation with Henderson Method 3 and MIVQUE0 techniques in a IBM 3081D mainframe computer with MVS/XA. Results obtained indicated that MIVQUE0 was more efficient than Method 3 in use of computer time and its efficiency increases with the level of total time used. The

differences between MIVQUEO and Method 3 in CPU time used were of 40 and 17,337 seconds when the total time needed to obtain the MIVQUEO estimates were of 942 and 24,721 seconds.

Effects of SOF misidentification on estimates of SOF variance components were studied using Henderson Method 1. It was found that as a result of SOF misidentification the expected values for error term increased. However, expected values for SOF variance components could not be obtained due to the existence of the fixed effect SOC in the model.

From this research, it can be concluded that SOF effects, even though they are real and measurable, do not have enough practical significance to be included in the mathematical models used in genetic analyses. The interaction effects SOF x HYS and SOF x SOC were measured and found to be real and measurable also; if their values are pooled together with those of HYS x SOC a new coefficient to replace the presently used c^2 would be obtained. This new c^2 is slightly smaller than 0.14 used by USDA in sires evaluation. Misidentification of SOF does not have any effect on the estimates of total variation. MIVQUEO is more efficient than Henderson Method 3 in use of computing time and its efficiency increases with the absolute time required for the estimations.

APPENDIX
PROOFS FOR FORMULAS IN TABLES 6.2 AND 6.3

1. $$E\left[\sum_j \frac{y_{\cdot j\cdot}^2}{n_{\cdot j}}\right] = E\left[\sum_j \frac{(\sum_i y_{ij\cdot})^2}{n_{\cdot j}}\right]$$

$$\sum_i y_{ij\cdot} = n_{\cdot j}\mu + \sum_i n_{ij}B_i + \sum_i n_{ij\ell}C_j + \sum_i (n_{ij}-n_{ij\ell})C_{jk} + \sum_i e_{ij\cdot}$$

$$E(\sum_i y_{ij\cdot})^2 = n_{\cdot j}^2\mu^2 + (\sum_i n_{ij}B_i)^2 + \sum_i n_{ij\ell}^2\sigma_c^2 + \sum_i (n_{ij}-n_{ij\ell})^2\sigma_c^2 + n_{\cdot j}\sigma_e^2$$

$$E\left[\sum_j \frac{(\sum_i y_{ij})^2}{n_{\cdot j}}\right] =$$

$$n_{\cdot\cdot}\mu^2 + \sum_{ij} n_{ij}B_i^2 + \frac{\sum_{ij} n_{ij}^2 + 2\sum_{ij} n_{ij\ell}(n_{ij\ell}-n_{ij})}{n_{\cdot\cdot}}\sigma_c^2 + r\sigma_e^2$$

2. $$E(CF) = E\left[\frac{y_{...}^2}{n..}\right] = E\left[\frac{(\sum_{ij} y_{ij.})^2}{n..}\right]$$

$$= E\left[\frac{(n..\mu+\sum_{ij} n_{ij}B_i+\sum_{ij} n_{ij\ell}C_j+\sum_{ij} (n_{ij}-n_{ij\ell})C_{jk}+\sum_{ij} e_{ij.})^2}{n..}\right]$$

$$= E\left[\frac{n_{..}^2\mu^2+(\sum_{ij} n_{ij}B_i)^2}{n..}\right] + E\left[\frac{(\sum_{ij} n_{ij\ell}C_j)^2+(\sum_{ij} (n_{ij}-n_{ij\ell})C_{jk})^2}{n..}\right]$$

$$+ E\left[\frac{(\sum_{ij} e_{ij.})^2}{n..}\right]$$

(By parts)

$$E\left[\frac{n_{..}^2\mu^2+(\sum_{ij} n_{ij}B_i)^2}{n..}\right] = n..\mu^2$$

$$E\left[\frac{(\sum_{ij} n_{ij\ell}C_j)^2+(\sum_{ij} (n_{ij}-n_{ij\ell})C_{jk})^2}{n..}\right] =$$

$$\frac{\sum_{ij} (n_{ij\ell}^2+(n_{ij}-n_{ij\ell})^2)}{n..}\sigma_c^2 = \frac{\sum_{ij} n_{ij}^2+2\sum_{ij} n_{ij\ell}(n_{ij\ell}-n_{ij})}{n..}\sigma_c^2$$

$$E\left[\frac{(\sum_{ij} e_{ij.})^2}{n..}\right] = \frac{n..\sigma_e^2}{n..} = \sigma_e^2$$

$$E(CF) = n..\mu^2 + \frac{\sum_{ij} n_{ij}^2+2\sum_{ij} n_{ij\ell}(n_{ij\ell}-n_{ij})}{n..}\sigma_c^2 + \sigma_e^2$$

3. $E(\text{Total SS}) = E(\sum_{ijm} y_{ijm}^2) = n..\mu^2+\sum_{ij} n_{ij}B_i^2+n..\sigma_c^2+n..\sigma_e^2$

4. $E\left[\sum_{ij} \frac{y_{ij\cdot}^2}{n_{ij}}\right]$

$$= E\left[\sum_{ij} \frac{n_{ij}^2\mu^2+n_{ij}^2B_i^2+n_{ij\ell}^2C_j^2+(n_{ij}-n_{ij\ell})^2C_{jk}^2+e_{ij\cdot}^2}{n_{ij}}\right]$$

$$= E\left[\sum_{ij} \frac{n_{ij}^2\mu^2+n_{ij}^2B_i^2}{n_{ij}}\right] + E\left[\sum_{ij} \frac{n_{ij\ell}^2C_j^2+(n_{ij}-n_{ij\ell})^2C_{jk}^2}{n_{ij}}\right]$$

$$+ E\left[\sum_{ij} \frac{e_{ij\cdot}^2}{n_{ij}}\right]$$

(By parts)

$$E\left[\sum_{ij} \frac{n_{ij}^2\mu^2+n_{ij}^2B_i^2}{n_{ij}}\right] = E[n..\mu^2+\sum_{ij} n_{ij}B_i^2] = n..\mu^2+\sum_{ij} n_{ij}B_i^2$$

$$E\left[\sum_{ij} \frac{n_{ij\ell}^2C_j^2+(n_{ij}-n_{ij\ell})^2C_{jk}^2}{n_{ij}}\right] = (n..+2\sum_{ij} \frac{n_{ij\ell}(n_{ij\ell}-n_{ij})}{n_{ij}})\sigma_c^2$$

$$E\left[\sum_{ij} \frac{e_{ij\cdot}^2}{n_{ij}}\right] = c\sigma_e^2$$

$$E\left[\sum_{ij} \frac{y_{ij\cdot}^2}{n_{ij}}\right] = n..\mu^2+\sum_{ij} n_{ij}B_i^2+ (n..+2\sum_{ij} \frac{n_{ij\ell}(n_{ij\ell}-n_{ij})}{n_{ij}})\sigma_c^2+c\sigma_e^2$$

$$5.\quad E(MSE) = E\left(\frac{SSE}{df}\right) = E\ \frac{\sum_{ijm} y^2_{ijm} - \sum_{ij} \frac{y^2_{ij\cdot}}{n_{ij}}}{df}$$

$$E\ \sum_{ijm} y^2_{ijm} - \sum_{ij} \frac{y^2_{ij\cdot}}{n_{ij}}$$

$$= n..\mu^2 + \sum_{ij} n_{ij}B_i^2 + n..\sigma_c^2 + n..\sigma_e^2 - \left[n..\mu^2 + \sum_{ij} n_{ij}B_i^2 + \right.$$

$$\left. \left(n.. + 2\sum_{ij} \frac{n_{ij\ell}(n_{ij\ell} - n_{ij})}{n_{ij}}\right) \sigma_c^2 + c\sigma_e^2\right]$$

(if $n_{ij} = n$, $n_{ij\ell} = pn$, $n.. = cn$, and $df = c(n-1)$)

$$= (cn-c)\sigma_e^2 - 2\sum_{ij=1}^{c} \frac{pn(pn-n)}{n}\ \sigma^2 = c(n-1)\sigma_e^2 - 2cpn(p-1)\sigma_c^2$$

(Dividing by $c(n-1)$)

$$E(\hat{\sigma}_e^2) = \sigma_e^2 - \frac{2pn(p-1)}{n-1}\ \sigma_c^2 = \sigma_e^2 + \frac{n(1-x)}{n-1}\ \sigma_c^2$$

BIBLIOGRAPHY

1. Adkinson, R.W. 1977. Genetic studies of production and reproduction of dairy cattle with especial emphasis on sire of fetus effects. Ph.D. Dissertation. Univ. Fla., Gainesville.

2. Adkinson, R.W., C.J. Wilcox, and W.W. Thatcher. 1977. Effect of sire of fetus upon subsequent production and days open of the dam. J. Dairy Sci. 60:1964.

3. Allaire, F.R., and S.N. Gaunt. 1965. First lactation contemporary comparisons as indicators of environmental influences on daughter records used for sire evaluation. J. Dairy Sci. 48:454.

4. Arora, K.K., and A.E. Freeman. 1971. Environmental correlation between paternal half-sisters for milk and milk fat production. J. Dairy Sci. 54:880.

5. Auran, T., E. Fimland, and H. Skjervold. 1977. The foetal effect on milk yield in current lactation. Paper presented at 28th Ann. Meet. of Europ. Assoc. Animal Prod., Brussels, August 22-25, 1977.

6. Balaine, D.S., and R.W. Touchberry. 1980. Effects of current service sire and the interaction of current service sire x breed of cow on yields of milk, fat, SNF and protein and on percentage composition of fat, protein and SNF. J. Animal Sci. 51, Suppl. 1:112 (Abstract).

7. Balaine, D.S., and R.W. Touchberry. 1980. Effects of previous service sire, sire and service sire x breed of cow interaction on percent milk component traits and days open. J. Dairy Sci. 63, Suppl. 1:96 (Abstract).

8. Bereskin, B., and A.E. Freeman. 1965. Genetic and environmental factors in dairy sire evaluation. II. Uses and limitations of deviation records and the role of dams. J. Dairy Sci. 48:352.

9. Bereskin, B., and J.L. Lush. 1963. Considerations with environmental correlations among daughters in bull proofs. Paper presented at the 58th Ann. Meet. of Am. Dairy Sci. Assoc., Lafayette, Indiana, June, 1963.

10. Bereskin, B., and J.L. Lush. 1965. Genetic and environmental factors in dairy sire evaluation. III. Influence of environmental and other extraneous correlations among the daughters. J. Dairy Sci. 48:356.

11. Burdick, J.M., and L.D. McGilliard. 1963. Interactions between sires in artificial insemination and management of dairy herds. J. Dairy Sci. 46:452.

12. Dairy Bulls in Artificial Insemination. 1982. National Association of Animal Breeders, P.O. Box 1033, Columbia, MO 65205.

13. Dickinson, F.N., H.D. Norman, J.F. Keown, and L.G. Waite. 1974. Revision to USDA methodology for sire summaries and cow indexes. J. Dairy Sci. 57:977.

14. Dickinson, F.N., H.D. Norman, R.I. Powell, L.G. Waite, and B.T. McDaniel. 1976. The USDA-DHIA modified contemporary comparison. Procedures used to calculate the USDA-DHIA modified contemporary comparison. Production Research Report No. 165. USDA, Beltsville, Maryland.

15. Eisenhart, C. 1947. The assumptions underlying analysis of variance. Biometrics 3:1.

16. Everett, R.W. 1974. An extension approach to new sire summaries. J. Dairy Sci. 57:972.

17. Falconer, D.S. 1983. Introduction to Quantitative Genetics. Longman Inc., New York.

18. Freeman, A.E. 1979. Components of variance: Their history, use, and problems in animal breeding. Variance Components and Animal Breeding. Ed. L.D. Van Vleck and S.R. Searle. Cornell University, Ithaca, New York.

19. Gaines, W.L. 1915. A contribution to the physiology of lactation. Am. J. Phisiol. 38:285.

20. Gaylor, D.W., H.L. Lucas, and B.L. Anderson. 1970. Calculation of expected mean squares by the Abbreviated Doolittle and Square Root Methods. Biometrics 26:641.

21. Goodnight, J.H. 1978. Computing MIVQUE0 estimates of variance components. SAS Technical Report R-105. SAS Institute Inc., Cary, North Carolina.

22. Goodnight, J.H., and F.M. Speed. 1978. Computing expected mean squares. SAS Technical Report R-102. SAS Institute Inc., Cary, North Carolina.

23. Hargrove, G.L., D.A. Mbah, and J.L. Rosenberger. 1981. Genetic and environmental influences on milk and milk component production. J. Dairy Sci. 64:1593.

24. Hartley, H.O. 1967. Expectations, variances and covariances of ANOVA mean squares by "synthesis." Biometrics 23:105.

25. Hartley, H.O., and J.N.K. Rao. 1967. Maximum likelihood estimation for the mixed analysis of variance model. Biometrika 54:99.

26. Hartley, H.O., J.N.K. Rao, and L.R. LaMotte. 1978. A simple "synthesis"-based method of variance component estimation. Biometrics 34:233.

27. Harvey, W.R. 1964. Computing procedures for a generalized least-squares analysis program. Paper presented at the NSF Analysis of Variance Conference. Colorado St. Univ., Ft. Collins, Colorado, July.

28. Harvey, W.R. 1970. Estimation of variance and covariance components in the mixed model. Biometrics 26:485.

29. Hayes, J.F., B.R. Mangurkar, and R.I. Cue. 1984. Effect of sire of foetus on production traits. J. Dairy Sci. 67, Suppl. 1:246 (Abstract).

30. Henderson, C.R. 1953. Estimation of variance and covariance components. Biometrics 9:226.

31. Henderson, C.R. 1956. Cornell research on methods of selecting dairy sires. New Zealand Soc. Animal Prod. Proc. 16:69.

32. Henderson, C.R. 1973. Sire evaluation and genetic trends. Proceedings of the animal breeding and genetics symposium in honor of Dr. Jay L. Lush. ASAS and ADSA, Champaigne, Illinois.

33. Henderson, C.R. 1974. General flexibility of linear model techniques for sire evaluation. J. Dairy Sci. 57:963.

34. Henderson, C.R. 1975. Comparison of alternative sire evaluation methods. J. Animal Sci. 41:760.

35. Johansson, I., and J. Rendel. 1968. Genetics and Animal Breeding. W.H. Freeman and Co., San Francisco.

36. Johnson, L.P., and L.D. Van Vleck. 1978. Components of variance for effects associated with sire of fetus and service sire on milk yield, gestation length, and days open. J. Dairy Sci. 61, Suppl. 1:87 (Abstract).

37. Johnson, L.P., and L.D. Van Vleck. 1979. Components of variance associated with service sire for milk yield and reproduction traits. J. Dairy Sci. 62:754.

38. Johnson, L.P., and L.D. Van Vleck. 1979. Estimation of the genetic correlation between direct and fetal effects on first lactations of daughters and mates of Holstein sires. J. Dairy Sci. 62, Suppl. 1:107 (Abstract).

39. Lee, A.J. 1976. Estimation of variance components in large herd-by-sire designs with interactions. J. Dairy Sci. 59:2138.

40. Legates, J.E., and F.J. Verlinder. 1956. Sire by herd interaction in production traits in dairy cattle. J. Dairy Sci. 39:1055.

41. Lin, C.Y., and A.J. McAllister. 1984. Monte Carlo comparison of four methods for estimation of genetic parameters in the univariate case. J. Dairy Sci. 67:2389.

42. Lindhe, B. 1980. Caesarean section and induced parturition in cattle. Anim. Breed. Abs. 48:2466.

43. Lindhe, B. 1980. The effect of the foetus on cow milk yield. Anim. Breed. Abs. 48:3866.

44. Lindstrom, U.B., J. Syvajarvi, and T. Bellman. 1981. Milk production of dairy cows inseminated with semen from beef or dairy bulls. Zeit. Tierz. Zuchtungsbiol. 98:303.

45. Lush, J.L. 1931. The number of daughters necessary to prove a sire. J. Dairy Sci. 14:209.

46. Lush, J.L. 1935. Progeny test and individual performance as indicators of an animal's breeding value. J. Dairy Sci. 18:1.

47. McDaniel, B.T. 1974. Why new summaries are needed. J. Dairy Sci. 57:951.

48. McDaniel, B.T., and J.E. Legates. 1965. Analysis of records expressed as deviations. J. Dairy Sci. 48:749.

49. Mbah, D.A. 1980. Genetic, environmental and pricing influences on milk and milk components production. Ph.D. Dissertation. Pennsylvania State Univ., University Park.

50. Norman, H.D. 1974. Factors that should be considered in a natural sire summary model. J. Dairy Sci. 57:955.

51. Norman, H.D., B.G. Cassell, F.N. Dickinson, and A.L. Kuck. 1979. USDA-DHIA milk components sire summary. Production Research Report No. 178. USDA, Beltsville, Maryland.

52. Norman, H.D., B.T. McDaniel, and F.N. Dickinson. 1972. Modified herdmate comparison procedure that weights deviation yields considering daughters' distribution across herds. J. Dairy Sci. 55:691 (Abstract).

53. Norman, H.D., B.T. McDaniel, and F.N. Dickinson. 1977. Conflicts between heritability estimates of mature equivalent and herdmate deviation milk and fat. J. Dairy Sci. 55:507.

54. Northeast Regional Data Center. 1982. NERDC User's Manual: Utilities. Univ. Fla., Gainesville, FL 32611.

55. Plowman, R.D., and B.T. McDaniel. 1968. Changes in USDA sire summary procedures. J. Dairy Sci. 51:306.

56. Quaas, R.L., and D.C. Bolgiano. 1979. Sampling variances of the MIVQUE and Method 3 estimators of the sire component of variance: A numerical comparison. Variance Components and Animal Breeding. Ed. L.D. Van Vleck and S.R. Searle. Cornell University, Ithaca, New York.

57. Quesnel, F.N. 1978. Effects of sire of fetus upon dairy cattle production and reproduction in a subtropical environment. Master's Thesis. Univ. Fla., Gainesville.

58. Quesnel, F.N., C.J. Wilcox, and W.W. Thatcher. 1979. Effects of sire of fetus upon subsequent production and reproduction of the dam. J. Dairy Sci. 62, Suppl. 1:216 (Abstract).

59. Rao, C.R. 1970. Estimation of heteroscedastic variances in linear models. J. Am. Stat. Assoc. 65:161.

60. Rao, C.R. 1971. Estimation of variance and covariance components--MINQUE theory. J. Multiv. Anal. 1:257.

61. Rao, C.R. 1971. Minimum variance quadratic unbiased estimation of variance components. J. Multiv. Anal. 1:445.

62. Rao, C.R. 1972. Estimation of variance and covariance components in linear models. J. Am. Stat. Assoc. 67:112.

63. Rao, J.N.K. 1968. On expectations, variances and covariances of ANOVA mean squares by "synthesis." Biometrics 24:963.

64. Rich, D.K. 1979. Estimation of variance components using residuals: Some empirical evidence. Variance Components and Animal Breeding. Ed. L.D. Van Vleck and S.R. Searle. Cornell University, Ithaca, New York.

65. SAS Institute Inc. 1980. SAS Applications Guide. SAS Institute Inc., Cary, North Carolina.

66. SAS Institute Inc. 1982. SAS Users Guide: Basics. SAS Institute Inc., Cary, North Carolina.

67. SAS Institute Inc. 1982. SAS Users Guide: Statistics. SAS Institute Inc., Cary, North Carolina.

68. Schaeffer, L.R. 1975. Disconnectedness and variance component estimation. Biometrics 31:969.

69. Searle, S.R. 1964. Review of sire-proving methods in New Zealand, Great Britain, and New York state. J. Dairy Sci. 47:402.

70. Searle, S.R. 1968. Another look at Henderson's methods of estimating variances components. Biometrics 24:749.

71. Searle, S.R. 1971. Linear Models. John Wiley & Sons, Inc., New York.

72. Searle, S.R. 1979. Maximum likelihood and minimum variance estimation of variance components. Variance Components and Animal Breeding. Ed. L.D. Van Vleck and S.R. Searle. Cornell University, Ithaca, New York.

73. Shannon, D. 1980. Sire-of fetus effects on performance of the dam's current calf in beef cattle. Ph.D. Dissertation. Univ. Tennessee, Knoxville.

74. Sharma, A.K., C.J. Wilcox, R.C. Littell, W.W. Thatcher, and R.W. Adkinson. 1979. MIVQUEO estimates of sire of fetus effects on milk yield and days open. J. Dairy Sci. 62, Suppl. 1:106.

75. Skjervold, H. 1979. Effect of foetal genotype on maternal performance. Indian J. Anim. Genet. and Breed. 1(1):23.

76. Skjervold, H., and E. Fimland. 1975. Evidence for a possible influence of the fetus on the milk yield of the dam. Zeit. Tierz. Zuchtungsbid. 92:245.

77. Speed, F.M., and R.R. Hocking. 1974. Computation of expectations, variances and covariances of ANOVA mean squares. Biometrics 30:157.

78. Taylor, St. C.S. 1978. The mate-effect story. Animal Breeding Research Organization Report. Agricultural Research Council. Edinburgh, Scotland.

79. Taylor, St. C.S., L.S. Monteiro, L.S. Murray, and T.J. Osmond. 1978. Possible association between the breeding value of dairy bulls and milk yield of their mates. Anim. Prod. 27:303.

80. Thomson, G.M., and A.E. Freeman. 1970. Environmental correlations in pedigree estimates of breeding value. J. Dairy Sci. 53:1259.

81. Touchberry, R.W., and D.S. Balaine. 1980. Effects of previous service sire, sire, and previous service sire x breed of cow interaction on milk, fat, SNF and protein yield. J. Dairy Sci. 63, Suppl. 1:96 (Abstract).

82. Van Vleck, L.D. 1966. Change in variance components associated with milk records with time and increase in mean production. J. Dairy Sci. 49:36.

83. Van Vleck, L.D. 1966. Environmental correlations in sire evaluation. J. Dairy Sci. 49:56.

84. Van Vleck, L.D. 1966. Paternal half-sib correlations between pairs in the same and different herds. J. Dairy Sci. 49:195.

85. Van Vleck, L.D. 1970. Misidentification in estimating the paternal sib correlation. J. Dairy Sci. 53:1469.

86. Van Vleck, L.D. 1970. Misidentification and sire evaluation. J. Dairy Sci. 53:1697.

87. Van Vleck, L.D. 1978. A genetic model involving fetal effects on traits of the dam. Biometrics 34:123.

88. Van Vleck, L.D. 1978. Economic weights for direct and fetal genetic effects in choosing sires. J. Dairy Sci. 61:970.

89. Van Vleck, L.D. 1979. Estimation of sire of fetus effects from DHI records. Report to Eastern Artificial Insemination Cooperative, Inc. Mimeo. Cornell University, Ithaca, New York.

90. Van Vleck, L.D., L.H. Wadell, and C.R. Henderson. 1961. Components of variance associated with milk and fat records of artificially sired Holstein daughters. J. Animal Sci. 20:812.

91. Verde, O.G., C.J. Wilcox, F.G. Martin, and C.W. Reaves. 1972. Sire by herd interaction for milk production in Florida DHIA herds. Bulletin 754. IFAS. Univ. Fla., Gainesville, Florida, December.

92. Wickham, B.W. 1978. Economic advantage of including mates production in a bull selection index. Unpublished. Cited by L.P. Johnson and L.D. Van Vleck. 1979. Components of variance associated with service sire for milk yield and reproduction traits. J. Dairy Sci. 62:754.

93. Yao, S.C., J.M. White, W.E. Vinson, and R.H. Kliewer. 1979. Environmental correlations among paternal half sisters for type. J. Dairy Sci. 62:493.

BIOGRAPHICAL SKETCH

Joaquim Moya was born on September 23, 1945, in Catalonia, Spain. In 1966, he moved to Peru where in 1971 he graduated from the Universidad Nacional Agraria with a B.S. in animal science. He worked for the Swiss Government Technical Aid for four years in the Amazon region as a cattle production specialist. Granted a scholarship from the Swiss Government, he came to the University of Florida and he graduated with a M.S. in dairy science in 1977. He then returned to work for the Swiss Government Technical Aid, this time as a leader in a large dairy project in the coastal Region of Peru. He came back to the University of Florida in 1981 as a graduate research assistant to pursue a Ph.D. degree. He has accepted a position after graduation with Ralston Purina International Co. in Barcelona, Spain.

Joaquim is married to Sharon S. Moya, Ph.D. candidate in Bilingual Education, and they have two daughters. He became a naturalized U.S. citizen in March, 1985.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Charles J. Wilcox, Chairman
Professor of Animal Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Robert J. Collier
Associate Professor of Dairy Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Frank G. Martin
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James R. Simpson
Professor of Food and Resource Economics

This dissertation was submitted to the Graduate Faculty of the College of Agriculture and to the Graduate School, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May, 1985

Dean, College of Agriculture

Dean, Graduate School